The Annals of Probability
2011, Vol. 39, No. 3, 1161–1203
DOI: 10.1214/10-AOP572

MERGING FOR INHOMOGENEOUS FINITE MARKOV CHAINS, PART II: NASH AND LOG-SOBOLEV INEQUALITIES

By L. Saloff-Coste[1] and J. Zúñiga[2]

Cornell University and Stanford University

We study time-inhomogeneous Markov chains with finite state spaces using Nash and logarithmic-Sobolev inequalities, and the notion of c-stability. We develop the basic theory of such functional inequalities in the time-inhomogeneous context and provide illustrating examples.

1. Introduction.

1.1. *Background.* This article is part of a series of works where we study quantitative merging properties of time inhomogeneous finite Markov chains. Time inhomogeneity leads to a great variety of behaviors. Moreover, even in rather simple situations, we are at a loss to study how a time inhomogeneous Markov chain might behave. Here, we focus on a natural but restricted type of problem. Consider a sequence of aperiodic irreducible Markov kernels $(K_i)_1^\infty$ on a finite set V. Let π_i be the invariant measure of K_i. Assume that, in a sense to be made precise, all K_i and all π_i are similar and the behavior of the time homogeneous chains driven by each K_i separately is understood. Can we then describe the behavior of the time inhomogeneous chain driven by the sequence $(K_i)_1^\infty$?

To give a concrete example, on $V_N = \{0, \dots, N\}$, consider a sequence of aperiodic irreducible birth and death chain kernels K_i, $i = 1, 2, \dots$, with

$$1/4 \le K_i(x, y) \le 3/4 \qquad \text{if } |x - y| \le 1$$

Received March 2009; revised April 2010.
[1]Supported in part by NSF Grant DMS-06-03886.
[2]Supported in part by NSF Grants DMS-06-03886, DMS-03-06194 and by a NSF Postdoctoral Fellowship.
MSC2010 subject classification. 60J10.
Key words and phrases. Time-inhomogeneous Markov chains, spectral techniques, Nash inequalities, log-Sobolev inequalities.

and with reversible measure π_i satisfying $1/4 \le (N+1)\pi_i(x) \le 4$, for all $x \in V_N$. What can we say about the behavior of the corresponding time inhomogeneous Markov chain?

Remarkably enough, there is very little known about this question. What can we expect to be true? What can we try to prove? Let $K_{0,n}(x,\cdot)$ denote the distribution, after n steps, of the time inhomogeneous chain described above started at x. It is not hard to see that such a chain satisfies a Doeblin type condition that implies

$$\lim_{n\to\infty} \|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} = 0.$$

In the absence of a true target distribution and following [4], we call this property *merging*. Of course, this does not qualify as a quantitative result. Extrapolating from the behavior of each kernel K_i taken individually, we may hope to show that, if $\lim_{N\to\infty} t_N/N^2 = \infty$ then

$$\lim_{N\to\infty} \|K_{0,t_N}(x,\cdot) - K_{0,t_N}(y,\cdot)\|_{\mathrm{TV}} = 0.$$

The aim of this paper and the companion paper [32] is to present techniques that apply to this type of problem. The simple minded problem outlined above is actually quite challenging and we will not be able to resolve it here without some additional hypotheses. However, we show how to adapt techniques such as singular values, Nash and log-Sobolev inequalities to time inhomogeneous chains and provide a variety of examples where these tools apply. In [32], we discussed singular value techniques. Here, we focus on Nash and log-Sobolev inequalities. The examples treated here (as well as those treated in [32, 33]) are quite particular despite the fact that one may believe that the techniques we use are widely applicable. Whether or not such a belief is warranted is a very interesting and, so far, unanswered question. This is deeply related to the notion of c-stability that is introduced here and in [32]. The examples we present here and in [30, 32, 33] are about the only existing evidence of successful quantitative analysis of time inhomogeneous Markov chains.

A more detailed introduction to these questions is in [32]. The references [17, 30] discuss singular value techniques in the case of time inhomogeneous chains that admit an invariant distribution [all kernels K_i in the sequence $(K_i)_1^\infty$ share a common invariant distribution]. Time inhomogeneous random walks on finite groups provide a large collection of such examples (see also [24] for a particularly interesting example: semirandom transpositions). The papers [7, 14] are also concerned with quantitative results for time inhomogeneous Markov chains. In particular, the techniques developed in [7] are closely related to ours and we will use some of their results concerning the modified logarithmic Sobolev inequality. References on the basic theory

of time inhomogeneous Markov chains are [19, 26, 35–37]. For a different perspective, see also [3].

A short review of the relevant aspects of the time inhomogeneous Markov chain literature, including the use of "ergodic coefficients" can be found in [34]. The vast literature on the famous simulated annealing algorithm is not very relevant for our purpose but we refer to [6] for a recent discussion. The paper [5] concerned with filtering and genetic algorithms describes problems that are related in spirit to the present work.

1.2. *Basic notation.* Let V be a finite set equipped with a sequence of kernels $(K_n)_1^\infty$ such that, for each n, $K_n(x,y) \geq 0$ and $\sum_y K_n(x,y) = 1$. An associated Markov chain is a V-valued random process $X = (X_n)_0^\infty$ such that, for all n,

$$\begin{aligned} P(X_n = y | X_{n-1} = x, \dots, X_0 = x_0) &= P(X_n = y | X_{n-1} = x) \\ &= K_n(x,y). \end{aligned}$$

The distribution μ_n of X_n is determined by the initial distribution μ_0 and given by

$$\mu_n(y) = \sum_{x \in V} \mu_0(x) K_{0,n}(x,y),$$

where $K_{n,m}(x,y)$ is defined inductively for each n and each $m \geq n$ by

$$K_{n,m}(x,y) = \sum_{z \in V} K_{n,m-1}(x,z) K_m(z,y)$$

with $K_{n,n} = I$ (the identity). If we interpret the K_n's as matrices, then this definition means that $K_{n,m} = K_{n+1} \cdots K_m$. This paper is mostly concerned with the behavior of the measures $K_{0,n}(x,\cdot)$ as n tends to infinity. In the case of time homogeneous chains where all $K_i = Q$ are equal, we write $K_{0,n} = Q^n$.

Our main interest is in ergodic like properties of time inhomogeneous Markov chains. In general, one does not expect $\mu_n = \mu_0 K_{0,n}$ to converge toward a limiting distribution. Instead, the natural notion is that of merging of measures as discussed in [4].

Definition 1.1. Fix a sequence of Markov kernels as above. We say the sequence is merging if for any $x, y, z \in V$,

$$\lim_{n \to \infty} K_{0,n}(x,z) - K_{0,n}(y,z) = 0. \tag{1.1}$$

Remark 1.2. If the sequence $(K_i)_1^\infty$ is merging then, for any two starting distributions μ_0, ν_0, the measures $\mu_n = \mu_0 K_{0,n}$ and $\nu_n = \nu_0 K_{0,n}$ are merging, that is, $\mu_n - \nu_n \to 0$. Since we assume the set V is finite, merging is equivalent to $\lim_{n\to\infty} \|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} = 0$. Hence, we also refer to this property as "total variation merging."

Total variation merging is also referred to as weak ergodicity in the literature and there exists a body of work concerned with understanding when weak ergodicity holds. See, for example, [19, 25–27, 35]. A main tool used to show weak ergodicity is that of contraction coefficients. Furthermore, in [16], Birkhoff's contraction coefficient is used to study *ratio ergodicity* which is equivalent to what we will later call relative-sup merging. However, it should be noted that even for time homogeneous chains Birkhoff coefficients and related methods fail to provide useful quantitative bounds in most cases.

Our goal is to develop quantitative results in the context of time inhomogeneous chains in the spirit of the work of Aldous, Diaconis and others. In these works, precise estimates of the mixing time of ergodic chains are obtained. Typically, a family of Markov chains indexed by a parameter, say N, is studied. Loosely speaking, as the parameter N increases, the complexity and size of the chain increases and one seeks bounds that depend on N in an explicit quantitative way. See, for example, [1, 2, 8–13, 15, 22, 23, 28]. Efforts in this direction for time inhomogeneous chains are in [7, 14, 16–18, 24, 30, 32]. Still, there are only a very small number of results and examples concerning the quantitative study of merging as defined above for time inhomogeneous Markov chains so that it is not very clear what kind of results should be expected and what kind of hypotheses are reasonable. We refer the reader to [32] for a more detailed discussion.

The following definition is useful to capture the spirit of our study. It indicates that the simplest case we would like to think about is the case when the sequence K_i is obtained by deterministic but arbitrary choices between a finite number of kernels $\mathcal{Q} = \{Q_1, \dots, Q_k\}$.

Definition 1.3. We say that a set $\mathcal{Q}$ of Markov kernels on V is merging in total variation if for any sequence $(K_i)_0^\infty$ with $K_i \in \mathcal{Q}$ for all i, we have

$$\forall x, y, z \in V \qquad \lim_{n\to\infty} \|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} = 0.$$

In the study of ergodicity of finite Markov chains, the convergence toward the target distribution is measured using various notions of distance between probability measures. These include the total variation distance

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{A\subset V}\{\mu(A) - \nu(A)\},$$

the chi-square distance (w.r.t. ν. Note the asymmetry between μ and ν.)

$$\left(\sum_y \left|\frac{\mu(y)}{\nu(y)} - 1\right|^2 \nu(y)\right)^{1/2},$$

and the relative sup-distance (again, note the asymmetry)

$$\max_y\left\{\left|\frac{\mu(y)}{\nu(y)} - 1\right|\right\}.$$

These will be used here to measure merging.

1.3. *Merging time.* In the quantitative theory of ergodic time homogeneous Markov chains, the notion of mixing time plays a crucial role. For time inhomogeneous chain, we propose to consider the following definitions.

DEFINITION 1.4. Fix $\varepsilon \in (0,1)$. Given a sequence $(K_i)_1^\infty$ of Markov kernels on a finite set V, we call max total variation merging time the quantity

$$T_{\mathrm{TV}}(\varepsilon) = \inf\Big\{n : \max_{x,y\in V} \|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} < \varepsilon\Big\}.$$

DEFINITION 1.5. Fix $\varepsilon \in (0,1)$. We say that a set $\mathcal{Q}$ of Markov kernels on V has max total variation ε-merging time at most T if for any sequence $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}$ for all i, we have $T_{\mathrm{TV}}(\varepsilon) \le T$, that is,

$$\forall t > T \qquad \max_{x,y\in V}\{\|K_{0,t}(x,\cdot) - K_{0,t}(y,\cdot)\|_{\mathrm{TV}}\} \le \varepsilon.$$

Of course, merging can be measured in ways other than total variation. Also merging is a bit less flexible than mixing in this respect since there is no reference measure. One very natural and much stronger notion than total variation is relative sup-distance. For time inhomogeneous chains, total variation merging does not necessarily imply relative-sup merging as defined below. See [32].

DEFINITION 1.6. We say a sequence $(K_i)_1^\infty$ of Markov kernels on a finite set V is merging in relative-sup if for all $x, y, z \in V$

$$\lim_{n\to\infty} \frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} = 1$$

with the convention that $0/0 = 1$ and $a/0 = \infty$ for $a > 0$. Fix $\varepsilon \in (0,1)$, we call relative-sup merging time the quantity

$$T_\infty(\varepsilon) = \inf\left\{n : \max_{x,y,z\in V}\left\{\left|\frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} - 1\right|\right\} < \varepsilon\right\}.$$

DEFINITION 1.7. We say a set $\mathcal{Q}$ of Markov kernels on V is merging in relative-sup if any sequence $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}$ for all i is merging in relative-sup.

Fix $\varepsilon \in (0,1)$. We say that $\mathcal{Q}$ has relative-sup ε-merging time at most T if for any sequence $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}$ for all i, we have $T_\infty(\varepsilon) \le T$, that is,

$$\forall t > T \qquad \max_{x,y,z\in V}\left\{\left|\frac{K_{0,t}(x,z)}{K_{0,t}(y,z)} - 1\right|\right\} \le \varepsilon.$$

The following problem is open. It is a quantitative version of the problem stated at the beginning of the introduction.

PROBLEM 1.8. Let $V_N = \{0,\dots,N\}$ and $c \in [1,\infty)$. Let $\mathcal{Q}_N$ be the set of all birth and death chains Q on V_N with $Q(x,y)\in[1/4,3/4]$ if $|x-y|\le 1$, and reversible measure π satisfying $1/4\le (N+1)\pi(x)\le 4$, $x\in V_N$.

1. Prove or disprove that there exists a constant A independent of N such that $\mathcal{Q}_N$ has total variation ε-merging time at most $AN^2(1+\log_+ 1/\varepsilon)$.
2. Prove or disprove that there exists a constant A independent of N such that $\mathcal{Q}_N$ has relative-sup ε-merging time at most $AN^2(1+\log_+ 1/\varepsilon)$.

REMARK 1.9. This problem is open (in most cases) even if one considers a sequence $(K_i)_1^\infty$ drawn from a set $\mathcal{Q}=\{K_1,K_2\}$ of two kernels. Observe that the hypothesis that the invariant measures π_i are all comparable to the uniform plays some role. How to harvest the global hypothesis of comparable stationary distributions π_i is not entirely clear. See Theorem 1.14 below for a partial solution.

If π_1 and π_2 are not comparable, it is possible for (K_1,π_1) and (K_2,π_2) to have the same mixing time yet for $\mathcal{Q}=\{K_1,K_2\}$ to have a merging time of a higher order. Assume that K_1 and K_2 are two biased random walks with equal drift, one drift to left, the other to the right. Despite the fact that each of these random walks has a relative-sup mixing time of order N, the inhomogeneous chain driven by the sequence $K_1K_2K_1K_2\cdots$ has a relative-sup merging time of order N^2, see [32].

1.4. *Stability.* In this section, we consider a property, c-stability, that plays a crucial role in the techniques we develop to provide quantitative bounds for time inhomogeneous Markov chains. This property was introduced and discussed in [32]. It is a straightforward generalization of the property of sharing the same invariant measure. Unfortunately, it is hard to check.

DEFINITION 1.10. Fix $c\ge 1$. A sequence of Markov kernels $(K_n)_1^\infty$ on a finite set V is c-stable if there exists a measure μ_0 such that

$$\forall n\ge 0, x\in V \qquad c^{-1}\le \frac{\mu_n(x)}{\mu_0(x)}\le c, \tag{1.2}$$

where $\mu_n=\mu_0K_{0,n}$. If this holds, we say that $(K_n)_1^\infty$ is c-stable with respect to the measure μ_0.

DEFINITION 1.11. A set $\mathcal{Q}$ of Markov kernels is c-stable with respect to a measure μ_0 if any sequence $(K_i)_1^\infty$ such that $K_i\in\mathcal{Q}$ for all i is c-stable with respect to μ_0.

REMARK 1.12. If all K_i share the same invariant distribution π then $(K_i)_1^\infty$ is 1-stable with respect to π.

REMARK 1.13. Suppose a set $\mathcal{Q}$ of aperiodic irreducible Markov kernels is c-stable with respect to a measure μ_0. Let π be an invariant measure for some $Q \in \mathcal{Q}$. Then we must have

$$x \in V, \qquad \frac{1}{c} \leq \frac{\pi(x)}{\mu_0(x)} \leq c.$$

Hence, $\mathcal{Q}$ is also c^2-stable with respect to π and any two invariant measures π, π' for kernels $Q, Q' \in \mathcal{Q}$ must satisfy

$$x \in V, \qquad \frac{1}{c^2} \leq \frac{\pi(x)}{\pi'(x)} \leq c^2.$$

The following theorem which relates to a special case of Problem 1.8 illustrates the role of c-stability.

THEOREM 1.14. *Let $V_N = \{0, \dots, N\}$. Let $\mathcal{Q}_N$ be the set of all birth and death chains Q on V_N with*

$$Q(x,y) \in [1/4, 3/4] \qquad \text{if } |x-y| \leq 1$$

and reversible measure π satisfying $1/4 \leq (N+1)\pi(x) \leq 4$, $x \in V_N$. Let $(K_i)_1^\infty$ be a sequence of birth and death Markov kernels on V_N with $K_i \in \mathcal{Q}_N$. Assume that $(K_i)_1^\infty$ is c-stable with respect to the uniform measure on V_N, for some constant $c \geq 1$ independent of N. Then there exists a constant $A = A(c)$ (in particular, independent of N) such that the relative-sup merging time for $(K_i)_1^\infty$ on V_N is bounded by

$$T_\infty(\varepsilon) \leq AN^2(1 + \log_+ 1/\varepsilon).$$

This will be proved later in a stronger form in Section 2.4. In [32] the weaker conclusion $T_\infty(\varepsilon) \leq AN^2(\log N + \log_+ 1/\varepsilon)$ was obtained using singular value techniques. Here, we will use Nash inequalities to obtain $T_\infty(\varepsilon) \leq AN^2(1 + \log_+ 1/\varepsilon)$.

It is possible that the set $\mathcal{Q}_N$ is c-stable with respect to the uniform measure for some c. Indeed, it is tempting to conjecture that this is the case although the evidence is rather limited (see also the discussion in [34]). If this is true, then Theorem 1.14 solves Problem 1.8. However, we do not know how to approach the problem of proving c-stability for $\mathcal{Q}_N$.

REMARK 1.15. While the assumption of c-stability in Theorem 1.14 is quite strong, Sections 4.2 and 5 of [32] give specific examples of families $\mathcal{Q}_N$ for which it holds. Further, we note that the question of whether or not c-stability holds is extremely natural and interesting in itself.

2. Singular values and Nash inequalities. One key idea in the study of Markov chains is to associate to a Markov kernel K the operator $K: f \mapsto Kf = \sum_y K(\cdot, y)f(y)$. In the case of time homogeneous chains, one uses the basic fact that this operator acts on $\ell^p(\pi)$ with norm 1 when π is an invariant measure.

In the case of time inhomogeneous chains, it is crucial to consider K as an operator between ℓ^p spaces with different measures in the domain and target spaces. The following simple observation is key.

Given a measure μ and a Markov kernel K on a finite set V, set $\mu' = \mu K$. Fix $p \in [1, \infty)$ and consider K as a linear operator

$$K = K_\mu : \ell^p(\mu') \to \ell^p(\mu), \qquad Kf(x) = \sum_y K(x, y)f(y). \tag{2.1}$$

Then

$$\|K\|_{\ell^p(\mu') \to \ell^p(\mu)} = \sup\{\|Kf\|_{\ell^p(\mu)} : f \in \ell^p(\mu'), \|f\|_{\ell^p(\mu')} \le 1\} = 1. \tag{2.2}$$

This follows from Jensen's inequality. See, for example, [7, 32]. We will use the notation K_μ whenever we need to emphasize the fact that K is viewed as an operator between $\ell^p(\mu K)$ and $\ell^q(\mu)$ for some $1 \le p, q \le \infty$. When the context is clear, we will drop the subscript μ as was done above.

2.1. *Using various distances.* Given a sequence of Markov kernels $(K_i)_1^\infty$, fix a starting measure μ_0 and set $\mu_n = \mu_0 K_{0,n}$. We will assume that $\mu_n > 0$ for all n. Note that if $\mu_0 > 0$ and K_n are all irreducible then $\mu_n > 0$ for all $n \ge 0$. We are interested in the behavior of

$$d_p(K_{0,n}(x, \cdot), \mu_n) = \left(\sum_y \left|\frac{K_{0,n}(x, y)}{\mu_n(y)} - 1\right|^p \mu_n(y)\right)^{1/p}, \qquad p \ge 1.$$

For $p \ge 1$, a classical argument involving the duality between ℓ^p and ℓ^q where $1 = 1/p + 1/q$, yields

$$d_p(K_{0,n}(x, \cdot), \mu_n) = \sup\left\{\left|\sum_y [K_{0,n}(x, y)f(y) - \mu_n(y)f(y)]\right| : \|f\|_{\ell^q(\mu_n)} \le 1\right\}$$

and one checks that the function

$$n \mapsto d_p(K_{0,n}(x, \cdot), \mu_n)$$

is nonincreasing (see [32]). Of course,

$$2\|K_{0,n}(x, \cdot) - \mu_n\|_{\mathrm{TV}} = d_1(K_{0,n}(x, \cdot), \mu_n)$$

and, if $1 \le p \le r \le \infty$,

$$d_p(K_{0,n}(x, \cdot), \mu_n) \le d_r(K_{0,n}(x, \cdot), \mu_n).$$

In particular,

$$2\|K_{0,n}(x,\cdot)-\mu_n\|_{\mathrm{TV}} \le d_2(K_{0,n}(x,\cdot),\mu_n) \tag{2.3}$$

and

$$\|K_{0,n}(x,\cdot)-K_{0,n}(y,\cdot)\|_{\mathrm{TV}} \le \max_{x\in V}\{d_2(K_{0,n}(x,\cdot),\mu_n)\}. \tag{2.4}$$

Further, if

$$\max_{x,z}\left\{\left|\frac{K_{0,n}(x,z)}{\mu_n(z)}-1\right|\right\} \le \varepsilon \le 1/2,$$

then

$$\max_{x,y,z}\left\{\left|\frac{K_{0,n}(x,z)}{K_{0,n}(y,z)}-1\right|\right\} \le 4\varepsilon.$$

To see the last inequality, note that if $1-\varepsilon \le a/b, c/b \le 1+\varepsilon$ with $\varepsilon \in (0,1/2)$ then

$$1-2\varepsilon \le \frac{1-\varepsilon}{1+\varepsilon} \le \frac{a}{c} \le \frac{1+\varepsilon}{1-\varepsilon} \le 1+4\varepsilon.$$

2.2. *Singular values.* In [32], we developed basic inequalities for $d_2(K_{0,n}(x,\cdot),\mu_n)$ based on singular value decompositions. The basic fact here is that, if μ is a probability measure on V, K a Markov kernel and $\mu' = \mu K$ then

$$d_2(K(x,\cdot),\mu')^2 = \sum_{i=1}^{|V|-1} |\psi_i(x)|^2\sigma_i^2,$$

where σ_i, $i = 0,\dots,|V|-1$, are the singular values of $K_\mu : \ell^2(\mu') \to \ell^2(\mu)$ in nonincreasing order, that is the square root of the eigenvalues of $K_\mu K_\mu^* : \ell^2(\mu) \to \ell^2(\mu)$ where $K_\mu^* : \ell^2(\mu) \to \ell^2(\mu')$ is the adjoint of $K_\mu : \ell^2(\mu') \to \ell^2(\mu)$. The ψ_i's form an orthonormal basis for $\ell^2(\mu)$ and are eigenfunctions of $K_\mu K_\mu^*$, ψ_i being associated with σ_i^2. Of course, the σ_i^2's can also be viewed as the eigenvalues of $K_\mu^* K_\mu : \ell^2(\mu') \to \ell^2(\mu')$.

In any case, a crucial fact for us here is that σ_1, the second largest singular value of $K_\mu : \ell^2(\mu') \to \ell^2(\mu)$, is also the norm of $K-\mu' = K_\mu - \mu' : \ell^2(\mu') \to \ell^2(\mu)$, that is,

$$\sup\{\|(K-\mu')f\|_{\ell^2(\mu)} : f \in \ell^2(\mu'), \|f\|_{\ell^2(\mu')} = 1\} = \sigma_1.$$

Given a sequence $(K_i)_1^\infty$ of Markov kernels on V and a positive measure μ_0, set $\mu_n = \mu_0 K_{0,n}$ and let $\sigma_1(K_i,\mu_{i-1})$ be the second largest singular value of $K_i : \ell^2(\mu_i) \to \ell^2(\mu_{i-1})$. Noting that

$$(K_{0,n}-\mu_n) = (K_1-\mu_1)(K_2-\mu_2)\cdots(K_n-\mu_n),$$

we obtain

$$\|K_{0,n} - \mu_n\|_{\ell^2(\mu_n)\to\ell^2(\mu_0)} \le \prod_1^n \sigma_1(K_i, \mu_{i-1}). \tag{2.5}$$

This inequality seems very promising and this is rather misleading. There is very little hope to compute or estimate the singular values $\sigma_i(K_i, \mu_{i-1})$, even if we have a good grasp on the kernel K_i. The reason is that $\sigma_1(K_i, \mu_{i-1})$ depends very much on the unknown measure μ_{i-1}. This is similar to the problem one faces when studying an irreducible aperiodic time homogeneous finite Markov chain for which one is not able to compute the stationary measure (although this case is rarely discussed, it is the typical case). For positive examples and a more detailed discussion, see [32].

2.3. *Dirichlet forms.* Given a reversible Markov kernel Q with reversible measure π on a finite set V, the associated Dirichlet form is

$$\begin{aligned}\mathcal{E}(f,f) = \mathcal{E}_{Q,\pi}(f,f) &= \langle (I-Q)f, f\rangle_\pi \\ &= \frac{1}{2}\sum_{x,y} |f(x) - f(y)|^2 \pi(x) Q(x,y).\end{aligned}$$

This definition is essential for the techniques considered in this paper. To illustrate this, we note that the singular value $\sigma_1(K_\mu, \mu)$ associated to a Markov kernel K and a positive probability measure μ is the square root of the second largest eigenvalue of $K_\mu^* K_\mu : \ell^2(\mu') \to \ell^2(\mu')$, $\mu' = \mu K$. This operator is associated with the Markov kernel

$$P(x,y) = \frac{1}{\mu'(x)} \sum_z \mu(z) K(z,x) K(z,y),$$

which is reversible with respect to μ' and has associated Dirichlet form

$$\mathcal{E}_{P,\mu'}(f,f) = \frac{1}{2}\sum_{x,y,z} |f(x) - f(y)|^2 \mu(z) K(z,x) K(z,y).$$

Hence, using the classical variational formula for eigenvalues, we have

$$1 - \sigma_1(K,\mu) = \inf\left\{ \frac{\mathcal{E}_{P,\mu'}(f,f)}{\mathrm{Var}_{\mu'}(f)} : f \in \ell^2(\mu'), \mathrm{Var}_{\mu'}(f) \neq 0 \right\},$$

where $\mathrm{Var}_{\mu'}(f) = \|f\|^2_{\ell^2(\mu')} - \mu'(f)^2 = \sum_x |f(x) - \mu'(f)|^2 \mu'(x)$.

2.4. *Nash inequalities.* The use of Nash inequalities to study the convergence of ergodic (time homogeneous) finite Markov chains was developed in [11] (Section 7 of [11] discusses time homogeneous chains that admits an invariant measure). We refer the reader to that paper for background on this technique. In this section, we observe that it can be implemented in the context of time inhomogeneous chains. We start with some basic material.

DEFINITION 2.1. Let V be a state space equipped with a Markov kernel K and probability measures μ and ν. If $1 \le p, q \le \infty$ then

$$\|K\|_{\ell^p(\mu)\to\ell^q(\nu)} = \sup_{\|f\|_{\ell^p(\mu)\le 1}} \{\|Kf\|_{\ell^q(\nu)}\}.$$

If p and q are conjugate exponents, that is, if $1/p + 1/q = 1$, then

$$\|f\|_{\ell^p(\mu)} = \sup_{\|g\|_{\ell^q(\mu)\le 1}} \{\langle f, g\rangle_\mu\}.$$

The following proposition is well known in a much more general context.

PROPOSITION 2.2. *Let K be a Markov kernel. Let $K_\mu : \ell^2(\mu K) \to \ell^2(\mu)$ be the Markov operator on V with adjoint $K_\mu^* : \ell^2(\mu) \to \ell^2(\mu K)$ with respect to the inner product*

$$\langle Kf, g\rangle_\mu = \langle f, K^* g\rangle_{\mu K}.$$

If $1 \le p, r, s \le \infty$, $1/p + 1/q = 1$ and $1/r + 1/s = 1$ then

$$\|K\|_{\ell^p(\mu K)\to\ell^r(\mu)} = \|K^*\|_{\ell^s(\mu)\to\ell^q(\mu K)}.$$

Let now $(K_i)_1^\infty$ be a sequence of Markov kernels on V. Fix a positive probability measure μ_0 and set $\mu_n = \mu_0 K_{0,n}$ as usual. Consider $K_i : \ell^2(\mu_i) \to \ell^2(\mu_{i-1})$, its adjoint $K_i^* : \ell^2(\mu_{i-1}) \to \ell^2(\mu_i)$ and $P_i = K_i^* K_i : \ell^2(\mu_i) \to \ell^2(\mu_i)$. The operator P_i is given by the Markov kernel

$$P_i(x,y) = \frac{1}{\mu_i(x)} \sum_z \mu_{i-1}(z) K_i(z,x) K_i(z,y). \tag{2.6}$$

This kernel is reversible with reversible measure μ_i. We let

$$\mathcal{E}_{P_i,\mu_i}(f,f) = \frac{1}{2} \sum_{x,y} |f(x) - f(y)|^2 \mu_i(x) P_i(x,y)$$

be the associated Dirichlet form on $\ell^2(\mu_i)$.

THEOREM 2.3. *Referring to the setup and notation introduced above, let $N \ge 1$ and assume that there are constants $C, D > 0$ such that for $1 \le m \le N$ the following Nash inequalities hold*

$$\begin{aligned} \forall f : V \to \mathbb{R} \qquad \|f\|_{\ell^2(\mu_m)}^{2+1/D} \le C\bigg(&\mathcal{E}_{P_m,\mu_m}(f,f) \\ &+ \frac{1}{N}\|f\|_{\ell^2(\mu_m)}^2\bigg)\|f\|_{\ell^1(\mu_m)}^{1/D}. \end{aligned} \tag{2.7}$$

Then, for $0 \le m \le n \le N$,

$$\max\{\|K_{m,n}\|_{\ell^2(\mu_n)\to\ell^\infty(\mu_m)}, \|K_{m,n}\|_{\ell^1(\mu_n)\to\ell^2(\mu_m)}\} \le \left(\frac{4CB}{n-m+1}\right)^D, \tag{2.8}$$

where $B = B(D,N) = (1+1/N)(1+\lceil 4D\rceil)$.

Proof. Let $(K_i)_0^\infty$ be a sequence of Markov kernels on V such that the Nash inequalities (2.7) hold. Pick a function f such that $\|f\|_{\ell^1(\mu_n)} = 1$. For $1 \le m \le n \le N$ define

$$t_n(n-m) = \|K_{m,n}f\|^2_{\ell^2(\mu_m)}.$$

Note that for any $n > 0$, $(t_n(i))_{i=0}^n$ is nonincreasing. Indeed, using the contraction property (2.2), we have

$$\begin{aligned} t_n(i+1) &= \|K_{n-i-1,n}f\|^2_{\ell^2(\mu_{n-i-1})} = \|K_{n-i}K_{n-i,n}f\|^2_{\ell^2(\mu_{n-i-1})} \\ &\le \|K_{n-i,n}f\|^2_{\ell^2(\mu_{n-i})} = t_n(i). \end{aligned}$$

Moreover, note that for any $0 \le i-1 \le n \le N$

$$t_n(i)^{1+1/(2D)} \le C(t_n(i) - t_n(i+1) + t_n(i)/N),$$

where C and D are the constants in (2.7). This follows by applying the Nash inequality to the function $K_{n-i,n}f$. Corollary 3.1 of [11] then yields that

$$t_n(i) \le \left(\frac{CB}{i+1}\right)^{2D}, \qquad 0 \le i \le n \le N,$$

where $B = B(D,N) = (1+1/N)(1+\lceil 4D\rceil)$. In particular, if $0 \le m \le n \le N$,

$$\|K_{m,n}\|_{\ell^1(\mu_n)\to\ell^2(\mu_m)} \le ((CB)/(n-m+1))^D.$$

From Proposition 2.2 it follows that, for $0 \le m \le n \le N$,

$$\|K^*_{m,n}\|_{\ell^2(\mu_m)\to\ell^\infty(\mu_n)} \le ((CB)/(n-m+1))^D.$$

Next we bound $\|K^*_{m,n}\|_{\ell^1(\mu_m)\to\ell^\infty(\mu_n)}$ for $0 \le m \le n \le N$. Consider the quantity $M(N)$ where

$$M(N) = \max_{0\le m\le n\le N}\{(n-m+1)^{2D}\|K^*_{m,n}\|_{\ell^1(\mu_m)\to\ell^\infty(\mu_n)}\}.$$

Let $l = \lfloor \frac{n-m}{2}\rfloor + m$, so that $0 \le m \le l \le n \le N$. We have

$$\begin{aligned} \|K^*_{m,n}\|_{\ell^1(\mu_m)\to\ell^\infty(\mu_n)} &\le \|K^*_{m,l}\|_{\ell^1(\mu_m)\to\ell^2(\mu_l)}\|K^*_{l,n}\|_{\ell^2(\mu_l)\to\ell^\infty(\mu_n)} \\ &\le \left(\frac{CB}{n-l+1}\right)^D \|K^*_{m,l}\|_{\ell^1(\mu_m)\to\ell^2(\mu_l)}. \end{aligned}$$

Note that for all $0 \le m \le l \le N$

$$\|K^*_{m,l}\|_{\ell^1(\mu_m)\to\ell^2(\mu_l)} \le \|K^*_{m,l}\|^{1/2}_{\ell^1(\mu_m)\to\ell^\infty(\mu_l)}\|K^*_{m,l}\|^{1/2}_{\ell^1(\mu_m)\to\ell^1(\mu_l)}. \tag{2.9}$$

This follows from the fact that for any function f

$$\|K^*_{m,l}f\|_{\ell^2(\mu_l)} \le \|K^*_{m,l}f\|^{1/2}_{\ell^\infty(\mu_l)}\|K^*_{m,l}f\|^{1/2}_{\ell^1(\mu_l)}.$$

By (2.2), we have

$$\begin{aligned}\|K^*_{m,n}\|_{\ell^1(\mu_m)\to\ell^\infty(\mu_n)} &\le \left(\frac{CB}{n-l+1}\right)^D \|K^*_{m,l}\|^{1/2}_{\ell^1(\mu_m)\to\ell^\infty(\mu_l)} \\ &\le \left(\frac{CB}{(n-l+1)(l-m+1)}\right)^D M(N)^{1/2} \\ &\le \left(\frac{4CB}{(n-m+1)^2}\right)^D M(N)^{1/2}.\end{aligned}$$

The last inequality follows from the fact that

$$n-l+1 \ge \frac{n-m+1}{2} \quad \text{and} \quad l-m+1 \ge \frac{n-m+1}{2}.$$

So we have $M(N) \le (4CB)^{2D}$ and it follows that for all $0 \le m \le n \le N$

$$\|K^*_{m,n}\|_{\ell^1(\mu_m)\to\ell^\infty(\mu_n)} \le \left(\frac{4CB}{n-m+1}\right)^{2D}.$$

By duality, we get that

$$\|K_{m,n}\|_{\ell^1(\mu_n)\to\ell^\infty(\mu_m)} \le \left(\frac{4CB}{n-m+1}\right)^{2D}.$$

Next, we use the Riesz–Thorin interpolation theorem, see [38], page 179, which gives us the desired result. □

The next results show how Theorem 2.3 together with the singular value technique of Section 2.2 yields merging results.

THEOREM 2.4. *Referring to the above setup and notation, let $N \ge 1$ and assume that there are constants $C, D > 0$ such that for $1 \le m \le N$ the Nash inequalities*

$$\begin{aligned}\forall f: V \to \mathbb{R} \qquad \|f\|^{2+1/D}_{\ell^2(\mu_m)} &\le C\left(\mathcal{E}_{P_m,\mu_m}(f,f) + \frac{1}{N}\|f\|^2_{\ell^2(\mu_m)}\right) \\ &\quad \times \|f\|^{1/D}_{\ell^1(\mu_m)}\end{aligned} \tag{2.10}$$

hold. Let $\sigma_1(K_m,\mu_{m-1})$ *be the second largest singular value of* $K_m:\ell^2(\mu_m)\to\ell^2(\mu_{m-1})$*, that is, the square root of the second largest eigenvalue of* P_m*. Then, for* $n>m$*,* $N\ge m\ge 0$*, we have*

$$d_2(K_{0,n}(x,\cdot),\mu_n)\le\left(\frac{8C(1+\lceil 4D\rceil)}{(m+1)}\right)^D\prod_{m+1}^n\sigma_1(K_i,\mu_{i-1}). \tag{2.11}$$

Moreover, for any $n=2m+u$*,* $0\le m\le N$*, we have*

$$\max_{x,y}\left\{\left|\frac{K_{0,n}(x,y)}{\mu_n(y)}-1\right|\right\}\le\left(\frac{8C(1+\lceil 4D\rceil)}{(m+1)}\right)^{2D}\prod_{m+1}^{m+u}\sigma_1(K_i,\mu_{i-1}). \tag{2.12}$$

Proof. We have

$$\max_{x\in V}\{d_2(K_{0,n}(x,\cdot),\mu_n)^2\}=\|K_{0,n}-\mu_n\|^2_{\ell^2(\mu_n)\to\ell^\infty(\mu_0)},$$

where μ_n is understood as the expectation operator $f\mapsto\mu_n(f)$. Moreover, for any $0\le m\le n$,

$$K_{0,n}-\mu_n=K_{0,m}(K_{m,n}-\mu_n),$$

because $K_{0,m}\mu_n f=K_{0,m}\mu_n(f)=\mu_n(f)$. Hence, for $0\le m\le N$,

$$\begin{aligned}d_2(K_{0,n}(x,\cdot),\mu_n)^2&\le\|K_{m,n}-\mu_n\|^2_{\ell^2(\mu_n)\to\ell^2(\mu_m)}\|K_{0,m}\|^2_{\ell^2(\mu_m)\to\ell^\infty(\mu_0)}\\&\le\left(\prod_{m+1}^n\sigma_1(K_i,\mu_{i-1})^2\right)\left(\frac{4CB}{m+1}\right)^{2D}.\end{aligned}$$

Using $B=N^{-1}(N+1)(1+\lceil 4D\rceil)$, gives (2.11). To obtain the stronger result (2.12), write

$$\max_{x,y\in V}\left\{\left|\frac{K_{0,n}(x,y)}{\mu_n(y)}-1\right|\right\}=\|K_{0,n}-\mu_n\|_{\ell^1(\mu_n)\to\ell^\infty(\mu_0)}$$

and

$$\begin{aligned}&\|K_{0,n}-\mu_n\|_{\ell^1(\mu_n)\to\ell^\infty(\mu_0)}\\&\quad\le\|K_{n-m,n}\|_{\ell^1(\mu_n)\to\ell^2(\mu_{n-m})}\times\|K_{m,n-m}-\mu_{n-m}\|_{\ell^2(\mu_{n-m})\to\ell^2(\mu_m)}\\&\qquad\times\|K_{0,m}\|_{\ell^2(\mu_m)\to\ell^\infty(\mu_0)}.\end{aligned}$$

The stated bound (2.12) follows. □

Just as we did for singular values, let us emphasize that the powerful looking results stated in this theorem are actually extremely difficult to apply. Again, the point is that the Dirichlet form $\mathcal{E}_{P_m,\mu_m}$, the space $\ell^2(\mu_m)$, and the singular values $\sigma_1(K_m,\mu_{m-1})$ all involve the unknown sequence of measures $\mu_n=\mu_0K_{0,n}$, $n=0,\dots$. The following subsection gives similar but more applicable results under additional hypotheses involving the notion of c-stability.

2.5. *Nash inequality under c-stability.* We state two results that parallel Theorems 5.9 and 5.10 of [32].

THEOREM 2.5. *Fix $c \in (1,\infty)$. Let $(K_i)_1^\infty$ be a sequence of irreducible Markov kernels on a finite set V. Assume that $(K_i)_1^\infty$ is c-stable with respect to a positive probability measure μ_0. For each i, set $\mu_0^i = \mu_0 K_i$ and let $\sigma(K_i,\mu_0)$ be the second largest singular value of $K_i = K_{i,\mu_0}$ as an operator from $\ell^2(\mu_0^i)$ to $\ell^2(\mu_0)$. Let $P_i^0 = K_{i,\mu_0}^* K_{i,\mu_0}$. Let $N \ge 1$ and assume that there are constants $C, D > 0$ such that for $1 \le m \le N$ the Nash inequalities*

$$\forall f : V \to \mathbb{R} \qquad \|f\|_{\ell^2(\mu_m^0)}^{2+1/D} \le C\left(\mathcal{E}_{P_m^0,\mu_m^0}(f,f) + \frac{1}{N}\|f\|_{\ell^2(\mu_m^0)}^2\right) \times \|f\|_{\ell^1(\mu_m^0)}^{1/D} \tag{2.13}$$

holds. Then, for $n > m$, $N \ge m \ge 0$, we have

$$d_2(K_{0,n}(x,\cdot),\mu_n) \le \left(\frac{8Cc^{2+3/2D}(1+\lceil 4D\rceil)}{(m+1)}\right)^D \times \prod_{m+1}^{n}\left(1 - \frac{1-\sigma(K_i,\mu_0)^2}{c^2}\right)^{1/2}. \tag{2.14}$$

Moreover, for any $n = 2m+u$, $0 \le m \le N$, we have

$$\max_{x,y}\left\{\left|\frac{K_{0,n}(x,y)}{\mu_n(y)} - 1\right|\right\} \le \left(\frac{8Cc^{2+3/2D}(1+\lceil 4D\rceil)}{(m+1)}\right)^{2D} \times \prod_{m+1}^{m+u}\left(1 - \frac{1-\sigma(K_i,\mu_0)^2}{c^2}\right)^{1/2}.$$

PROOF. First note that since $\mu_{i-1}/\mu_0 \in [1/c, c]$, we have $\mu_0^i/\mu_i \in [1/c, c]$. Consider the operator P_i with kernel

$$P_i(x,y) = \frac{1}{\mu_i(x)}\sum_z \mu_{i-1}(z)K_i(z,x)K_i(z,y).$$

By assumption

$$\mu_i(x)P_i(x,y) \ge c^{-1}\mu_0^i(x)\left[\frac{1}{\mu_0^i(x)}\sum_z \mu_0(z)K_i(z,x)K_i(z,y)\right],$$

where the term in brackets on the right-hand side is the kernel of P_i^0. This kernel has second largest eigenvalue $\sigma(K_i,\mu_0)^2$. A simple eigenvalue comparison argument yields

$$1 - \sigma_1(K_i,\mu_{i-1})^2 \ge \frac{1}{c^2}(1 - \sigma(K_i,\mu_0)^2).$$

Further, comparison of measures and Dirichlet form yields the Nash inequality

$$\forall f : V \to \mathbb{R} \qquad \|f\|_{\ell^2(\mu_m)}^{2+1/D} \le C c^{2+3/2D} \left(\mathcal{E}_{P_m,\mu_m}(f,f) + \frac{1}{N} \|f\|_{\ell^2(\mu_m)}^2 \right) \times \|f\|_{\ell^1(\mu_m)}^{1/D}.$$

Together with Theorem 2.4, this gives the stated result. □

The next result is based on a stronger hypothesis.

THEOREM 2.6. *Fix $c \in (1, \infty)$. Let $\mathcal{Q}$ be a family of irreducible aperiodic Markov kernels on a finite set V. Assume that $\mathcal{Q}$ is c-stable with respect to some positive probability measure μ_0.*

Let $(K_i)_1^\infty$ be a sequence of Markov kernels with $K_i \in \mathcal{Q}$ for all i. Let π_i be the invariant measure of K_i. Let $\tilde{P}_i = K_i^* K_i$ where $K_i : \ell^2(\pi_i) \to \ell^2(\pi_i)$. Let $\sigma_1(K_i)$ be the second largest singular value of K_i as an operator on $\ell^2(\pi_i)$. Let $N \ge 1$ and assume that there are constants $C, D > 0$ such that for $1 \le m \le N$ the Nash inequalities

$$\forall f : V \to \mathbb{R} \qquad \|f\|_{\ell^2(\pi_m)}^{2+1/D} \le C \left(\mathcal{E}_{\tilde{P}_m,\pi_m}(f,f) + \frac{1}{N} \|f\|_{\ell^2(\pi_m)}^2 \right) \times \|f\|_{\ell^1(\pi_m)}^{1/D}. \tag{2.15}$$

Then, for $n > m$, $N \ge m \ge 0$, we have

$$d_2(K_{0,n}(x,\cdot),\mu_n) \le \left(\frac{8 C c^{4+3/D}(1 + \lceil 4D \rceil)}{(m+1)} \right)^D \times \prod_{m+1}^{n} \left(1 - \frac{1 - \sigma_1(K_i)^2}{c^4} \right)^{1/2}. \tag{2.16}$$

Moreover, for any $n = 2m + u$, $0 \le m \le N$, we have

$$\max_{x,y} \left\{ \left| \frac{K_{0,n}(x,y)}{\mu_n(y)} - 1 \right| \right\} \le \left(\frac{8 C c^{4+3/D}(1 + \lceil 4D \rceil)}{(m+1)} \right)^{2D} \prod_{m+1}^{m+u} \left(1 - \frac{1 - \sigma_1(K_i)^2}{c^4} \right)^{1/2}.$$

PROOF. Note that the hypothesis that $\mathcal{Q}$ is c-stable implies $\pi_i/\mu_j \in [1/c^2, c^2]$ for all i, j. Consider again the operator P_i and its kernel

$$P_i(x,y) = \frac{1}{\mu_i(x)} \sum_z \mu_{i-1}(z) K_i(z,x) K_i(z,y).$$

By assumption

$$\begin{aligned}\mu_i(x)P_i(x,y) &\geq c^{-2}\pi_i(x)\left[\frac{1}{\pi_i(x)}\sum_z \pi_i(z)K_i(z,x)K_i(z,y)\right]\\ &\geq c^{-2}\pi_i(x)\tilde{P}_i(x,y).\end{aligned}$$

A comparison argument similar to the one used in the previous proof yields the desired result. □

3. Examples involving Nash inequalities. This section describes applications of the Nash inequality technique to several examples. All these examples are of the following general type.

(1) There is a basic reversible model (K,π) on a space V_N (growing with N) that is well understood because:

- We have good grasp on the second largest singular value σ_N of (K,π).
- The model (K,π) satisfies a good Nash inequality, that is, an inequality of the form

$$\|f\|_{\ell^2(\pi)}^{2+1/D} \leq BT_N\left(\mathcal{E}_{K^*K,\pi}(f,f) + \frac{1}{bT_N}\|f\|_{\ell^2(\pi)}^2\right)\|f\|_{\ell^1(\pi)}^{1/D}$$

  with B, b independent of N and $T_N \simeq (1-\sigma_N)^{-1}$. Here, $f \simeq g$ implies that there exist constants $d, D > 0$ such that $dg \leq f \leq Dg$.
- Together, the Nash inequality and second largest singular value estimate yield the mixing time estimate

$$\max_{x,y}\left\{\left|\frac{K^t(x,y)}{\pi(y)} - 1\right|\right\} \leq \eta, \qquad t \geq \frac{A(1+\log_+ 1/\eta)}{1-\sigma_N},$$

  where A is independent of N.

(2) We are given a sequence $(K_i)_1^\infty$ or a set $\mathcal{Q}_N$ of Markov kernels on V_N which satisfies:

- $(K_i)_1^\infty$ or $\mathcal{Q}_N$ is c-stable with respect to a measure μ_0 which is either equal or at least comparable to π.
- The Markov kernels K_i or the elements of $\mathcal{Q}_N$ are all bounded perturbations of K in the sense that $K_i(x,y)/K(x,y)$ is bounded away from 0 and away from ∞ for all $(x,y) \in V_N^2$. In particular, $K_i(x,y) = 0$ if and only if $K(x,y) = 0$.

Under such circumstances, Theorem 2.5 (or Theorem 2.6) applies and yields the conclusion that the time inhomogeneous Markov chain associated with the sequence K_i under investigation has a relative-sup merging time $T_\infty(\eta)$ bounded by

$$T_\infty(\eta) \leq \frac{A'(1+\log_+ 1/\eta)}{1-\sigma_N}$$

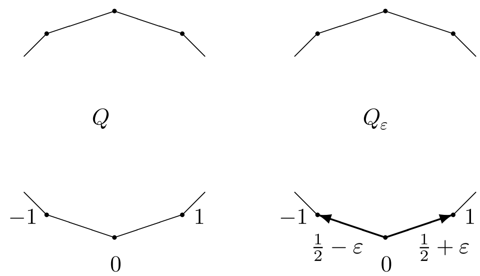

FIG. 1. *The asymmetric perturbation.*

for some constant A' independent of N.

The most obvious basic model is, perhaps, the simple random walk on $\mathbb{Z}/N\mathbb{Z}$ (with some holding if N is even to avoid periodicity). This model has $1-\sigma_N \simeq 1/N^2$ and satisfies the desired Nash inequality with $D=1/4$. The first subsection presents applications to a perturbation of this model.

3.1. *Asymmetric perturbation at the middle vertex.* In this example, $V_N = \mathbb{Z}/p_N\mathbb{Z}$ is a finite circle. It will be convenient to enumerate the points in V_N by writing $V_N = \{-(N-1),\dots,-1,0,1,\dots,(N-1),N\}$ if $p_N = 2N$ and $V_N = \{-N,\dots,-1,0,1,\dots,N\}$ if $p_N = 2N+1$. The simple random walk in V has kernel

$$Q(x,y) = \begin{cases} 1/2, & \text{if } |x-y| = 1, \\ 0, & \text{otherwise,} \end{cases} \tag{3.1}$$

and reversible measure $u \equiv \frac{1}{p_N}$. For any $\varepsilon > 0$, define the perturbation kernel

$$\Delta_\varepsilon(x,y) = \begin{cases} \varepsilon, & \text{if } (x,y) = (0,1), \\ -\varepsilon, & \text{if } (x,y) = (0,-1), \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

For $\varepsilon \in (-1/2, 1/2)$, the Markov kernel $Q_\varepsilon = Q + \Delta_\varepsilon$ is a perturbation of Q. See Figure 1.

For any fixed $0 < \varepsilon < 1/2$, set

$$\mathcal{Q}(\varepsilon) = \{Q_\delta : \delta \in [-\varepsilon, \varepsilon]\}.$$

We shall see below that $\mathcal{Q}(\varepsilon)$ is c-stable.

DEFINITION 3.1. Let $\mathcal{S}_N(\varepsilon)$ be the set of all probability measures on V_N which satisfy the following two properties:

(1) for all $x \in V_N$, there exist constants $a_{\mu,x}$ such that $a_{\mu,x} = -a_{\mu,-x}$ and

$$\mu(x) = (1/p_N) + a_{\mu,x}$$

(2) for all $x \in V_N$ we have that $|a_{\mu,x}| \leq 2\varepsilon/p_N$.

REMARK 3.2. Note that we always have $a_{\mu,0} = 0$ (since $-0 = 0$) and, in the case when $p_N = 2N$, $a_{\mu,N} = 0$.

CLAIM 3.3. *Let $\mu \in \mathcal{S}_N(\varepsilon)$ defined above, then for any $K \in \mathcal{Q}(\varepsilon)$ we have that $\mu K \in \mathcal{S}_N(\varepsilon)$.*

PROOF. Let $\mu \in \mathcal{S}_N(\varepsilon)$ and $K = Q_\delta \in \mathcal{Q}(\varepsilon)$, $\delta \in [-\varepsilon, \varepsilon]$. We show that μK has the properties required to be in $\mathcal{S}_N(\varepsilon)$.

(1) Any measure $\mu \in \mathcal{S}_N$ can be written as $\mu = u + m_\mu$ where m_μ is the (nonprobability) measure $m_\mu(x) = a_{\mu,x}$. A simple calculation yields that

$$m_\mu Q(x) = (a_{\mu,x-1} + a_{\mu,x+1})/2.$$

Since $a_{\mu,x} = -a_{\mu,-x}$, we obtain that

$$m_\mu Q(x) = -m_\mu Q(-x) \quad \text{and} \quad m_\mu Q(0) = 0.$$

The fact that $\mu Q = (u + m_\mu)Q = u + m_\mu Q$ implies that μQ satisfies property (1) in the definition of $\mathcal{S}_N(\varepsilon)$. To see that $\mu Q_\delta \in \mathcal{S}_N(\varepsilon)$ also satisfies this property, we note that

$$\mu\Delta_\delta(x) = \begin{cases} \delta\mu(0), & \text{if } x = 1, \\ -\delta\mu(0), & \text{if } x = -1, \\ 0, & \text{otherwise.} \end{cases}$$

It now follows that $\mu Q_\delta \in \mathcal{S}_N$ has property (1) in the definition of $\mathcal{S}_N(\varepsilon)$ since $\mu Q_\delta = \mu(Q + \Delta_\delta)$.

(2) We consider the measure μK. For $x \notin \{-1, 1\}$ property (2) of $\mathcal{S}_N(\varepsilon)$ follows easily from the fact that $|a_{\mu,x}| \leq 2\varepsilon/p_N$ and

$$\mu K(x) = 1/p_N + 1/2(a_{\mu,x-1} + a_{\mu,x+1}).$$

For $x = 1$, we note that

$$\begin{aligned} \mu K(1) &\leq \mu(0)(1/2 + \varepsilon) + \mu(2)(1/2) = 1/p_N + \varepsilon/p_N + (1/2)a_{\mu,2} \\ &\leq 1/p_N + 2\varepsilon/p_N. \end{aligned}$$

Similarly

$$\begin{aligned} \mu K(1) &\geq \mu(0)(1/2 - \varepsilon) + \mu(2)(1/2) = 1/p_N - \varepsilon/p_N - (1/2)a_{\mu,2} \\ &\geq 1/p_N - 2\varepsilon/p_N. \end{aligned}$$

The proof now follows from the fact that $a_{\mu K,1} = -a_{\mu K,-1}$ as proved in part (1) above. $\square$

Claim 3.4. *The family $\mathcal{Q}(\varepsilon)$ is $\frac{1+2\varepsilon}{1-2\varepsilon}$-stable with respect to any $\mu_0 \in \mathcal{S}_N(\varepsilon)$.*

Proof. Claim 3.3 implies that for any sequence $(K_i)_0^\infty$ such that $K_i \in \mathcal{Q}_\varepsilon$ and any measure $\mu_0 \in \mathcal{S}_N(\varepsilon)$ we have $\mu_n = \mu_0 K_{0,n} \in \mathcal{S}_N(\varepsilon)$ for all $n \geq 0$. Note that for any measure $\nu \in \mathcal{S}_N(\varepsilon)$ we have that

$$\nu(x) = 1/p_N + a_{\nu,x} \leq (1+2\varepsilon)/p_N \quad \text{and} \quad \nu(x) = 1/p_N + a_{\nu,x} \geq (1-2\varepsilon)/p_N.$$

Hence,

$$\frac{1-2\varepsilon}{1+2\varepsilon} \leq \frac{\mu_n(x)}{\mu_0(x)} \leq \frac{1+2\varepsilon}{1-2\varepsilon}. \qquad \square$$

When $p_N = 2N$, the kernels Q_δ yield periodic chains on V_N. In this case, we will study the merging properties of

$$\mathcal{Q}_{\mathrm{lazy}}(\varepsilon) = \{\tfrac{1}{2}(I+K) : K \in \mathcal{Q}(\varepsilon)\},$$

that is, the so-called lazy version of $\mathcal{Q}(\varepsilon)$. We set

$$\overline{Q}_\delta = \tfrac{1}{2}(I + Q_\delta).$$

For any $\mu \in \mathcal{S}_N(\varepsilon)$, we consider the kernel

$$P_{\delta,\mu}(x,y) = \frac{1}{\mu\overline{Q}_\delta(x)} \sum_z \mu(z)\overline{Q}_\delta(z,x)\overline{Q}_\delta(z,y),$$

which is the kernel of K^*K where $K = \overline{Q}_\delta : \ell^2(\mu\overline{Q}_\delta) \to \ell^2(\mu)$. This is 0 unless $y = x, x \pm 1, x \pm 2$ and we compare it to

$$\begin{aligned} P(x,y) = P_{0,u}(x,y) &= \frac{1}{u(x)} \sum_z u(z)\overline{Q}(z,x)\overline{Q}(z,y) \\ &= \sum_z \overline{Q}(z,x)\overline{Q}(z,y), \end{aligned}$$

which is $3/8$ if $y = x$, $1/4$ if $y = x \pm 1$, $1/16$ if $y = x \pm 2$ and 0 otherwise. The definitions of Q_δ and $\mathcal{S}_N(\varepsilon)$ yield

$$\begin{aligned} \mu\overline{Q}_\delta(x) P_{\delta,\mu}(x,y) &\geq \frac{(1-2\varepsilon)(1-2\delta)^2}{(1+2\varepsilon)} u(x)P(x,y) \\ &\geq \frac{(1-2\varepsilon)^3}{(1+2\varepsilon)} u(x)P(x,y). \end{aligned}$$

This yields

$$\mathcal{E}_{P,\mu}(f,f) \leq \frac{(1+2\varepsilon)}{(1-2\varepsilon)^3} \mathcal{E}_{P_{\delta,\mu},\mu\overline{Q}_\delta}(f,f), \tag{3.3}$$

whereas the stability property implies that the relevant measures $\mu\overline{Q}_\delta$ and u satisfy

$$\frac{(1-2\varepsilon)}{(1+2\varepsilon)}u \le \mu\overline{Q}_\delta \le \frac{(1+2\varepsilon)}{(1-2\varepsilon)}u. \tag{3.4}$$

In the case when $p_N = 2N+1$, we may work directly with the kernels Q_δ as they are not periodic. An analysis similar to that above will give versions of (3.3) and (3.4) for Q_δ.

Applying the line of reasoning explained at the beginning of this section and using Theorem 2.6, we get the following result.

THEOREM 3.5. *Fix $\varepsilon \in (0, 1/2)$. For any $\eta > 0$ the total variation η-merging time of the family $\mathcal{Q}_{\text{lazy}}(\varepsilon)$ on $V_N = \mathbb{Z}/2N\mathbb{Z}$ [resp., $\mathcal{Q}(\varepsilon)$ on $V_N = \mathbb{Z}/(2N+1)\mathbb{Z}$] is at most $B(\varepsilon)N^2(1+\log_+ 1/\eta)$ for some constant $B(\varepsilon) \in (0,\infty)$. In fact, we can choose $B(\varepsilon)$ such that*

$$\forall n \ge B(\varepsilon)N^2(1+\log_+ 1/\eta) \qquad \max_{x,y\in V_N}\left\{\left|\frac{K_{0,n}(x,z)}{K_{0,n}(y,z)} - 1\right|\right\} \le \eta$$

for any sequence $K_i \in \mathcal{Q}_{\text{lazy}}(\varepsilon)$ [resp., $K_i \in \mathcal{Q}(\varepsilon)$].

3.2. *Perturbations of some birth and death chains.* In [29], Nash inequalities are used to study certain birth and death chains on $V_N = \{-N, \dots, 0, \dots, N\}$ with reversible measures which belong to one of the following two families:

$$\hat{\pi}_\alpha(x) = \hat{c}(\alpha, N)(N - |x| + 1)^\alpha, \qquad \alpha \ge 0,$$

and

$$\check{\pi}_\alpha(x) = \check{c}(\alpha, N)(|x|+1)^\alpha, \qquad \alpha \ge 0.$$

Here, we consider $\alpha \in [0,\infty)$ to be a fixed parameter and are interested in what happens when N tends to infinity. From this perspective, the normalizing constants $\hat{c}(\alpha, N), \check{c}(\alpha, N)$ are comparable and behave as

$$\hat{c}(\alpha, N) \simeq \check{c}(\alpha, N) \simeq N^{-\alpha-1}.$$

Set

$$\zeta(\alpha, N) = \sum_0^N (1+i)^{-\alpha} \simeq \begin{cases} 1, & \text{if } \alpha > 1, \\ \log N, & \text{if } \alpha = 1, \\ N^{-\alpha+1}, & \text{if } \alpha \in [0,1). \end{cases}$$

Here, all $\simeq$ must be understood for fixed α and the implied comparison constants depend on α. Let $\hat{M}_\alpha$ (resp., $\check{M}_\alpha$) be the Markov kernel of the Metropolis chain with basis the symmetric simple random walk on V_N with holding $1/3$ at all points except at the end points where the holding is

2/3, and target $\hat{\pi}_\alpha$, (resp., $\check{\pi}_\alpha$). Let $\hat{\lambda}(\alpha, N)$, $\check{\lambda}(\alpha, N)$ be the corresponding spectral gaps. Let $\hat{T}(\alpha, N, \eta)$, $\check{T}(\alpha, N, \eta)$ be the relative-sup mixing times of these chains. It is proved in [29] that

$$\hat{\lambda}(\alpha, N) \simeq 1/N^2, \qquad \hat{T}(\alpha, N, \eta) \simeq N^2(1 + \log_+ 1/\eta),$$

whereas

$$\begin{aligned} \check{\lambda}(\alpha, N) &\simeq \check{c}(\alpha, N)/\zeta(\alpha, N), \\ \check{T}(\alpha, N, \eta) &\simeq (N^2 + [\check{c}(\alpha, N)/\zeta(\alpha, N)] \log_+ 1/\eta). \end{aligned}$$

Note that

$$\check{c}(\alpha, N)/\zeta(\alpha, N) \simeq \begin{cases} N^{-(1+\alpha)}, & \text{if } \alpha > 1, \\ (N^2 \log N)^{-1}, & \text{if } \alpha = 1, \\ N^{-2}, & \text{if } \alpha \in [0, 1). \end{cases}$$

These results are based on the Nash inequalities satisfied by these chains. Namely, letting $\mathcal{E}_\alpha = \mathcal{E}_{\hat{M}_\alpha, \hat{\pi}_\alpha}$ or $\mathcal{E}_\alpha = \mathcal{E}_{\check{M}_\alpha, \check{\pi}_\alpha}$ and $\pi_\alpha = \hat{\pi}_\alpha$ or $\pi_\alpha = \check{\pi}_\alpha$, there are constants $A_\alpha, a_\alpha \in (0, \infty)$ such that

$$\|f\|_{\ell^2(\pi_\alpha)}^{2+1/D_\alpha} \le A_\alpha N^2 \left(\mathcal{E}_\alpha(f, f) + \frac{1}{a_\alpha N^2} \|f\|^2_{\ell^2(\pi_\alpha)} \right) \|f\|_{\ell^1(\pi_\alpha)}^{1/D_\alpha}$$

with $D_\alpha = 1 + \alpha$. See [29].

In cite [32], the authors consider the class of birth and death chains Q on $V_N = \{-N, \dots, 0, \dots, N\}$ that are symmetric with respect to the middle point, that is, satisfy $Q(x, x+1) = Q(-x, -x-1)$, $Q(x, x-1) = Q(-x, -x+1)$, $Q(x, x) = Q(-x, -x)$, $x \in \{0, N\}$. For any such chain Q, let ν be the reversible measure. It satisfies $\nu(x) = \nu(-x)$. Consider the perturbation set

$$\mathcal{Q}_N(Q, \varepsilon) = \{Q + \Delta_s : s \in [-\varepsilon, \varepsilon]\}, \qquad \varepsilon \in [0, q_0),$$

where $q_0 = Q(0, \pm 1)$, $\Delta_s(0, \pm 1) = \pm s$ and $\Delta(x, y) = 0$ otherwise. These perturbations at the middle vertex have reversible measure ν_s that satisfy

$$\nu_s(0) = \nu(0), \qquad \nu_s(\pm x) = \nu(\pm x)(1 \pm s/q_0), \qquad x \in \{1, \dots, N\}.$$

The main point of this construction is the following.

PROPOSITION 3.6. *Fix Q, ν as above and $\varepsilon \in [0, q_0)$. The set $\mathcal{Q}_N(Q, \varepsilon)$ is c-stable with respect to $\mu_0 = \nu$ with $c = (q_0 + \varepsilon)/(q_0 - \varepsilon)$.*

In order to apply this results to our example $\hat{M}_\alpha, \check{M}_\alpha$, we observe that

$$\hat{q}_0(\alpha) = \hat{M}_\alpha(0, -1) = \frac{1}{3} \left(\frac{N}{N+1} \right)^\alpha$$

and

$$\check{q}_0(\alpha) = \check{M}_\alpha(0, -1) = \tfrac{1}{3}.$$

Now, Theorem 2.6 yields the following result.

THEOREM 3.7. *Fix* $\alpha \in [0, \infty)$ *and set* $\hat{\varepsilon}_{N,\alpha} = \frac{1}{6}(N/(N+1))^{\alpha}$, $\check{\varepsilon}_{N,\alpha} = 1/6$.

1. *There exists a constant* A *independent of* N *such that, for any sequence* $(K_i)_1^\infty$ *with* $K_i \in \mathcal{Q}_N(\hat{M}_\alpha, \hat{\varepsilon}_{N,\alpha})$, *we have*
$$T_\infty(\eta) \le AN^2(1 + \log_+ 1/\eta).$$
2. *There exists a constant* A *independent of* N *such that, for any sequence* $(K_i)_1^\infty$ *with* $K_i \in \mathcal{Q}_N(\check{M}_\alpha, \check{\varepsilon}_{N,\alpha})$, *we have*
$$T_\infty(\eta) \le A \begin{cases} N^2 + N^{1+\alpha}\log_+ 1/\eta, & \text{if } \alpha > 1, \\ N^2 + (N^2 \log N)\log_+ 1/\eta, & \text{if } \alpha = 1, \\ N^2(1 + \log_+ 1/\eta), & \text{if } \alpha \in (0,1). \end{cases}$$

4. Logarithmic Sobolev inequalities. This section develops the technique of logarithmic Sobolev inequality for time inhomogeneous finite Markov chains. It should be noted that the logarithmic Sobolev technique has been mostly applied in the literature in the context of continuous time chains. In [21], Miclo tackled the problem of adapting this technique to discrete time (time homogeneous) chains. There are two different ways to use logarithmic Sobolev inequality for mixing estimates. One, the most powerful, provides results for relative-sup merging and is based on hypercontractivity. The other is based on entropy and only produces bounds for total variation merging. We will discuss and illustrate both approaches below in the context of time inhomogeneous chains. The entropy approach is already treated in [7].

4.1. *Hypercontractivity.* Recall that, for any positive probability distribution μ, a Markov kernel K can be thought of as a contraction
$$K_\mu : \ell^2(\mu') \to \ell^2(\mu) \qquad \text{for } \mu' = \mu K.$$
The adjoint $K_\mu^* : \ell^2(\mu) \to \ell^2(\mu')$ has kernel
$$K_\mu^*(x,y) = \frac{K(y,x)\mu(y)}{\mu'(x)}.$$
Set $P = K_\mu^* K_\mu : \ell^2(\mu') \to \ell^2(\mu')$. We define the logarithmic Sobolev constant
$$l(P) = \inf\left\{ \frac{\mathcal{E}_{P,\mu'}(f,f)}{\mathcal{L}(f^2,\mu')} : \mathcal{L}(f^2,\mu') \ne 0, f \ne \text{constant} \right\},$$
where the ℓ^2 relative entropy $\mathcal{L}(f^2,\nu)$ of a function f with respect to the measure ν is defined by
$$\mathcal{L}(f^2,\nu) = \sum_{x\in V} f^2 \log\left(\frac{f^2}{\|f\|^2_{\ell^2(\nu)}}\right)\nu(x).$$

The following proposition is a slight generalization of [21], Proposition 2, in that it allows for the necessary change of measure.

PROPOSITION 4.1. *Let K and μ be a Markov kernel and a probability measure, respectively. For all $q_0 \geq 2$ and $q \leq [1 + l(P)]q_0$, then*

$$\|K\|_{\ell^{q_0}(\mu') \to \ell^q(\mu)} \leq 1.$$

In order to prove the proposition above, we will need the following two lemmas from [21].

LEMMA 4.2 ([21], Lemma 3). *Let ν be a probability measure. For all $q \geq q_0 \geq 1$,*

$$\|f\|_{\ell^q(\nu)} - \|f\|_{\ell^{q_0}(\nu)} \leq \frac{q - q_0}{q_0 q} \|f\|_{\ell^q(\nu)}^{1-q} \mathcal{L}(f^{q/2}, \nu).$$

LEMMA 4.3 ([21], Lemma 4). *Fix $\nu \geq 0$ and $q \geq 2$, then for any $t \geq 0$ and $-t \leq s \leq \nu t$ we have that*

$$(t + s)^q \geq t^q + qt^{q-1}s + g(q, \nu)((t + s)^{q/2} - t^{q/2})^2,$$

where

$$g(q, v) = \frac{(1 + \nu)^q - 1 - q\nu}{((1 + \nu)^{q/2} - 1)^2}.$$

The proof of Proposition 4.1 follows directly that of Proposition 2 in [21].

PROOF OF PROPOSITION 4.1. To prove Proposition 4.1 is suffices to only consider positive functions. For $f > 0$, we begin by writing

$$\begin{aligned} \|Kf\|_{\ell^q(\mu)} - \|f\|_{\ell^2(\mu')} = {} & \|Kf\|_{\ell^q(\mu)} - \|f\|_{\ell^q(\mu')} \\ & + \|f\|_{\ell^q(\mu')} - \|f\|_{\ell^2(\mu')}. \end{aligned} \tag{4.1}$$

The difference of the last two terms on the right-hand side is controlled by Lemma 4.2. To control the first two terms, we will use the concavity result

$$\forall a, b \geq 0 \qquad a^{1/q} - b^{1/q} \leq \frac{1}{q} b^{1/q-1}(a - b).$$

It follows that

$$\|Kf\|_{\ell^q(\mu)} - \|f\|_{\ell^q(\mu')} \leq \frac{1}{q} \|f\|_{\ell^q(\mu')}^{1-q} (\|Kf\|_{\ell^q(\mu)}^q - \|f\|_{\ell^q(\mu')}^q).$$

Set

$$\nu(K) = \max\{1/K(x, y) : K(x, y) > 0\} - 1.$$

Following the notation of Lemma 4.3, fix $x, y \in V$ and set $\nu = \nu(K)$, $t = Kf(x)$ and $t + s = f(y)$. If $K(x, y) > 0$, then $-t \leq s \leq \nu t$ and so

$$\begin{aligned} f(y)^q \geq\ & Kf(x)^q + qKf(x)^{q-1}(f(y) - Kf(x)) \\ & + g(q, \nu(K))(f(y)^{q/2} - Kf(x)^{q/2})^2. \end{aligned}$$

Fix x and integrate with respect to the measure $K(x, \cdot)$ to get

$$Kf^q(x) \geq (Kf(x))^q + g(q, \nu(K)) \sum_{y \in V} K(x, y)(f(y)^{q/2} - Kf(x)^{q/2})^2.$$

We also have

$$\begin{aligned} \sum_{y \in V} K(x, y)(f^{q/2}(y) - (Kf(x))^{q/2})^2 &\geq \min_{c \in \mathbb{R}} \sum_{y \in V} K(x, y)(f^{q/2}(y) - c)^2 \\ &= \sum_{y \in V} K(x, y)(f^{q/2}(y) - K(f^{q/2})(x))^2 \\ &= Kf^q(x) - (Kf^{q/2}(x))^2. \end{aligned}$$

Hence,

$$Kf^q(x) \geq (Kf(x))^q + g(q, \nu(K))(Kf^q(x) - (Kf^{q/2}(x))^2).$$

Integrating with respect to μ gives us that

$$\|f\|^q_{\ell^q(\mu')} \geq \|Kf\|^q_{\ell^q(\mu)} + g(q, \nu(K))\mathcal{E}_{P,\mu'}(f^{q/2}, f^{q/2}). \tag{4.2}$$

It follows from Lemma 4.2, (4.1) and (4.2) that

$$\begin{aligned} & \|Kf\|_{\ell^q(\mu)} - \|f\|_{\ell^2(\mu')} \\ & \quad \leq \frac{1}{q}\|f\|^{1-q}_{\ell^q(\mu')}\left(\frac{q - q_0}{q_0}\mathcal{L}(f^{q/2}, \mu') - g(q, \nu(K))\mathcal{E}_{P,\mu'}(f^{q/2}, f^{q/2})\right). \end{aligned}$$

In [21], it is noted that for all $\nu > 0$ and $q \geq 2$ we have $g(q, \nu) \geq 1$. So if $q \leq [1 + l(P)]q_0$ then $q \leq [1 + g(q, \nu(K))l(P)]q_0$. Hence,

$$\begin{aligned} & \|Kf\|_{\ell^q(\mu)} - \|f\|_{\ell^2(\mu')} \\ & \quad \leq \frac{1}{q}\|f\|^{1-q}_{\ell^q(\mu')} g(q, \nu(K))(l(P)\mathcal{L}(f^{q/2}, \mu') - \mathcal{E}_{P,\mu'}(f^{q/2}, f^{q/2})). \end{aligned}$$

Since $l(P)$ is the logarithmic Sobolev constant, we get our desired result,

$$\|Kf\|_{\ell^q(\mu)} - \|f\|_{\ell^2(\mu')} \leq 0. \qquad \square$$

Corollary 4.4. *Let $(K_n)_0^\infty$ be a sequence of Markov kernels on a finite set V and μ_0 be an initial distribution on V. Set $\mu_n = \mu_0 K_{0,n}$. Consider*

$K_i:\ell^2(\mu_i)\to\ell^2(\mu_{i-1})$ and $P_i=K_i^*K_i:\ell^2(\mu_i)\to\ell^2(\mu_i)$. Let $l(P_i)$ be the logarithmic Sobolev constant of P_i. Then for any $q_0\ge 2$ and $q\le\prod_{i=1}^n(1+l(P_i))q_0$, we have that

$$\|K_{0,n}\|_{\ell^{q_0}(\mu_n)\to\ell^q(\mu_0)}\le 1.$$

Proof. When $n=2$, set $q_1=(1+l(P_2))q_0$, then $q=(1+l(P_1))q_1$. It follows from Proposition 4.1 that

$$\|K_{0,2}\|_{\ell^{q_0}(\mu_2)\to\ell^q(\mu_0)}\le\|K_2\|_{\ell^{q_0}(\mu_2)\to\ell^{q_1}(\mu_1)}\|K_1\|_{\ell^{q_1}(\mu_1)\to\ell^{q_2}(\mu_0)}\le 1.$$

The proof by induction follows similarly. □

We now relate the results above to bounds on merging times.

Theorem 4.5. *Let V be a finite set equipped with a sequence of Markov kernels $(K_n)_0^\infty$ and an initial distribution μ_0. Let $\mu_n=\mu_0K_{0,n}$. Consider $K_i:\ell^2(\mu_i)\to\ell^2(\mu_{i-1})$ and $P_i=K_i^*K_i:\ell^2(\mu_i)\to\ell^2(\mu_i)$. Let $l(P_i)$ be the logarithmic Sobolev constant of P_i. Set*

$$m_x=\min\left\{t\in\mathbb{N}:\sum_{i=1}^t\log(1+l(P_i))\ge\log\log(\mu_0(x)^{-1/2})\right\}.$$

Then for $n\ge m_x$, we have that

$$d_2(K_{0,n}(x,\cdot),\mu_n)^2\le e^2\prod_{i=m_x+1}^n\sigma_1(K_i,\mu_{i-1})^2.$$

Proof. Fix x, and let $m=m_x$. If $0\le m\le n$, $K_{0,n}^*=K_{m,n}^*K_{0,m}^*$. Indeed, for any $f\in\ell^2(\mu_0)$ and $g\in\ell^2(\mu_n)$ we have that

$$\langle K_{0,n}^*f,g\rangle_{\mu_n}=\langle f,K_{0,n}g\rangle_{\mu_0}=\langle K_{0,m}^*f,K_{m,n}g\rangle_{\mu_m}=\langle K_{m,n}^*K_{0,m}^*f,g\rangle_{\mu_n}.$$

Moreover, if μ_m is thought of as the expectation operator $\mu_m:\ell^2(\mu_m)\to\ell^2(\mu_n)$, $f\mapsto\mu_m(f)$, then $(K_{m,n}^*-\mu_m)^*=K_{m,n}-\mu_n$. Let

$$\delta_x(z)=\begin{cases}\mu_0(x)^{-1}, & \text{if } z=x,\\ 0, & \text{otherwise.}\end{cases}$$

Set $q=q(m)=2\prod_{i=1}^m(1+l(P_i))$ and $q'(m)$ to be the conjugate exponent of $q(m)$ so that $1/q(m)+1/q'(m)=1$. By duality, we have

$$\begin{aligned}&d_2(K_{0,n}(x,\cdot),\mu_n)\\&\quad=\left\|\frac{K_{0,n}(x,\cdot)}{\mu_n(\cdot)}-1\right\|_{\ell^2(\mu_n)}=\left\|\frac{K_{0,n}^*(\cdot,x)}{\mu_0(x)}-1\right\|_{\ell^2(\mu_n)}\end{aligned}$$

$$
\begin{aligned}
&= \|(K_{0,n}^* - \mu_0)\delta_x\|_{\ell^2(\mu_n)} = \|(K_{m,n}^* - \mu_m)K_{0,m}^*\delta_x\|_{\ell^2(\mu_n)} \\
&\le \|K_{m,n}^* - \mu_m\|_{\ell^2(\mu_m)\to\ell^2(\mu_n)} \|K_{0,m}^*\delta_x\|_{\ell^2(\mu_m)} \\
&\le \|\delta_x\|_{\ell^{q'(m)}(\mu_0)} \|K_{0,m}^*\|_{\ell^{q'(m)}(\mu_0)\to\ell^2(\mu_m)} \|K_{m,n}^* - \mu_m\|_{\ell^2(\mu_m)\to\ell^2(\mu_n)} \\
&\le \mu_0(x)^{-1/q(m)} \|K_{0,m}\|_{\ell^2(\mu_m)\to\ell^{q(m)}(\mu_0)} \|K_{m,n} - \mu_n\|_{\ell^2(\mu_n)\to\ell^2(\mu_m)}.
\end{aligned}
$$

By assumption, we have that $q(m) \ge \log(\mu_0(x)^{-1})$, it now follows from Corollary 4.4 that

$$
d_2(K_{0,n}(x,\cdot),\mu_n) \le e \prod_{i=m+1}^{n} \sigma_1(K_i,\mu_{i-1}). \qquad \square
$$

4.2. *Logarithmic Sobolev inequalities and c-stability.*

THEOREM 4.6. *Fix $c \in (1,\infty)$. Let V be a finite set equipped with a sequence of irreducible Markov kernels, $(K_i)_1^\infty$. Assume that $(K_i)_1^\infty$ is c-stable with respect to a positive probability measure μ_0. For each i, set $\mu_0^i = \mu_0 K_i$ and let $\sigma_1(K_i,\mu_0)$ be the second largest singular value of the operator $K_i : \ell^2(\mu_0^i) \to \ell^2(\mu_0)$ and $l(K_i^* K_i)$ the logarithmic Sobolev constant for the operator $K_i^* K_i : \ell^2(\mu_0^i) \to \ell^2(\mu_0^i)$. If*

$$
\tilde{m}_x = \min\left\{ t \in \mathbb{N} : \sum_{i=1}^{t} \log(1 + c^{-2} l(K_i^* K_i)) \ge \log\log(\mu_0(x)^{-1/2}) \right\},
$$

then for $n \ge \tilde{m}_x$ we have that

$$
d_2(K_{0,n}(x,\cdot),\mu_n)^2 \le e^2 \prod_{i=\tilde{m}_x+1}^{n} (1 - c^{-2}(1 - \sigma_1(K_i,\mu_0^i)^2)).
$$

PROOF. First, we note that $\mu_i/\mu_0^i \in [c^{-1}, c]$. Let P_i be the Markov kernel described in the statement of Theorem 4.5. By the same arguments as in Theorem 2.5, we get that for all $x, y \in V$

$$
\mu_i(x) P_i(x,y) = \sum_z \mu_{i-1}(z) K_i(z,x) K_i(z,y) \ge c^{-1} \mu_0^i(x) K_i^* K_i(x,y).
$$

A simple comparison argument similar to those used in the proof of Theorem 2.5 (see also [10, 12]) yields that

$$
l(P_i) \ge c^{-2} l(K_i^* K_i) \quad \text{and} \quad 1 - \sigma(K_i,\mu_{i-1})^2 \ge c^{-2}(1 - \sigma(K_i,\mu_0^i)^2).
$$

The first inequality implies that $\tilde{m}_x \ge m_x$ where m_x is defined in the proof of Theorem 4.5. Using the results of Theorem 4.5 and the second inequality above gives the desired result. $\square$

The next result is when we have a c-stability assumption on a family of kernels.

THEOREM 4.7. *Let $c\in(1,\infty)$. Let $\mathcal{Q}$ be a family of irreducible aperiodic Markov kernels on a finite set V. Assume that $\mathcal{Q}$ is c-stable with respect to some positive probability measure μ_0. Let $(K_i)_1^\infty$ be a sequence of Markov kernels with $K_i\in\mathcal{Q}$ for all i. Let π_i be the invariant measure of K_i. Let $\sigma_i(K_i)$ be the second largest singular value for the operator $K_i:\ell^2(\pi)\to\ell^2(\pi)$. Let $l(K_i^*K_i)$ be the logarithmic Sobolev constant for the operator $K_i^*K_i$ where K_i^* is the adjoint of $K_i:\ell^2(\pi)\to\ell^2(\pi)$. If*

$$\tilde{m}_x=\min\left\{t\in\mathbb{N}:\sum_{i=1}^{t}\log(1+c^{-4}l(K_i^*K_i))\geq\log\log(\mu_0(x)^{-1/2})\right\},$$

then for $n\geq\tilde{m}_x$ we have that

$$d_2(K_{0,n}(x,\cdot),\mu_n)^2\leq e^2\prod_{i=m_x+1}^{n}(1-c^{-4}(1-\sigma_1(K_i)^2)).$$

PROOF. Let $\mu_i=\mu_0K_{0,i}$. If $\mathcal{Q}$ is c-stable, then $\mu_i/\pi_i\in[c^{-2},c^2]$. Similar arguments to those used in Theorem 4.6 give the desired result. $\square$

4.3. *The relative sup norm.* To control the relative-sup merging time by this method, we need an additional hypothesis. In the case of the ℓ^2 distance, we only required a control over the logarithmic Sobolev constant of the kernel $P_i=K_i^*K_i:\ell^2(\mu_i)\to\ell^2(\mu_i)$. In this case, we will also need to control the logarithmic Sobolev constant of $\check{P}_i=K_iK_i^*:\ell^2(\mu_{i-1})\to\ell^2(\mu_{i-1})$ where K_i^* is the adjoint of the operator K_i from $\ell^2(\mu_i)$ to $\ell^2(\mu_{i-1})$.

THEOREM 4.8. *Let V be a finite set equipped with a sequence of Markov kernels $(K_n)_0^\infty$ and an initial distribution μ_0. Let $\mu_n=\mu_0K_{0,n}$ and $P_i=K_i^*K_i:\ell^2(\mu_i)\to\ell^2(\mu_i)$ and $\check{P}_i=K_iK_i^*:\ell^2(\mu_{i-1})\to\ell^2(\mu_{i-1})$ where K_i^* is the adjoint of K_i with respect to the measure μ_i. Let $l(P_i)$ and $l(\check{P}_i)$ be the logarithmic Sobolev constants of P_i and $\check{P}_i$, respectively. If $\mu_i^\#=\min_x\{\mu_i(x)\}$ and*

$$m_0^\#=\min\left\{t\in\mathbb{N}:\sum_{i=1}^{t}\log(1+l(P_i))\geq\log\log(\mu_0^{\#-1/2})\right\},$$

$$m_n^\#=\min\left\{t\in\mathbb{N}:\sum_{i=n-t}^{n}\log(1+l(\check{P}_i))\geq\log\log(\mu_n^{\#-1/2})\right\},$$

then for any $n\geq 2m$,

$$\max_{x,y}\left\{\left|\frac{K_{0,n}(x,y)}{\mu_n(y)}-1\right|\right\}\leq e^2\prod_{i=m+1}^{n-m}\sigma_1(K_i,\mu_{i-1}),$$

where $m=\max\{m_0^\#,m_n^\#\}$.

REMARK 4.9. This innocent looking theorem is not easy to apply. For instance, m depends on n and without some control on this dependence the result is useless.

PROOF OF THEOREM 4.8. Write

$$\max_{x,y}\left\{\left|\frac{K_{0,n}(x,y)}{\mu_n(y)}-1\right|\right\}=\|K_{0,n}-\mu_n\|_{\ell^1(\mu_n)\to\ell^\infty(\mu_0)}$$

and

$$\begin{aligned}&\|K_{0,n}-\mu_n\|_{\ell^1(\mu_n)\to\ell^\infty(\mu_0)}\\&\quad\le\|K_{n-m,n}-\mu_n\|_{\ell^1(\mu_n)\to\ell^2(\mu_{n-m})}\times\|K_{m,n-m}-\mu_{n-m}\|_{\ell^2(\mu_{n-m})\to\ell^2(\mu_m)}\\&\qquad\times\|K_{0,m}-\mu_m\|_{\ell^2(\mu_m)\to\ell^\infty(\mu_0)}.\end{aligned}$$

Note that

$$\|K_{m,n-m}-\mu_{n-m}\|_{\ell^2(\mu_{n-m})\to\ell^2(\mu_m)}\le\prod_{i=m+1}^{n-m}\sigma_1(K_{i,\mu_{i-1}})$$

so we just need to bound the remaining terms in the right-hand side of the inequality above. To bound $\|K_{n-m,n}-\mu_n\|_{\ell^1(\mu_n)\to\ell^2(\mu_{n-m})}$ set $q^*=q^*(m)=2\prod_{i=1}^m(1+l(\check{P}_{n-m+i}))$ and write

$$\begin{aligned}&\|K_{n-m,n}-\mu_n\|_{\ell^1(\mu_n)\to\ell^2(\mu_{n-m})}\\&\quad=\|K^*_{n-m,n}-\mu_{n-m}\|_{\ell^2(\mu_{n-m})\to\ell^\infty(\mu_n)}\\&\quad=\|I(K^*_{n-m,n}-\mu_{n-m})\|_{\ell^2(\mu_{n-m})\to\ell^\infty(\mu_n)}\\&\quad\le\|K^*_{n-m,n}-\mu_{n-m}\|_{\ell^2(\mu_{n-m})\to\ell^{q^*}(\mu_n)}\|I\|_{\ell^{q^*}(\mu_n)\to\ell^\infty(\mu_n)}\\&\quad\le\|K^*_{n-m,n}\|_{\ell^2(\mu_{n-m})\to\ell^{q^*}(\mu_n)}\|I\|_{\ell^{q^*}(\mu_n)\to\ell^\infty(\mu_n)}.\end{aligned}$$

It follows from Corollary 4.4 that

$$\|K_{n-m,n}-\mu_n\|_{\ell^1(\mu_n)\to\ell^2(\mu_{n-m})}\le\|I\|_{\ell^{q^*}(\mu_n)\to\ell^\infty(\mu_n)}\le\mu_n^{\#-1/q^*}.$$

By assumption, we have that $q^*=q^*(m)\ge\log(\mu_n^{\#-1})$ so we get

$$\|K_{n-m,n}-\mu_n\|_{\ell^1(\mu_n)\to\ell^2(\mu_{n-m})}\le e.$$

To bound $\|K_{0,m}-\mu_0\|_{\ell^1(\mu_m)\to\ell^2(\mu_0)}$ set $q=q(m)=2\prod_{i=1}^m(1+l(P_i))$ and write

$$\|K_{0,m}-\mu_0\|_{\ell^2(\mu_m)\to\ell^\infty(\mu_0)}\le\|K_{0,m}-\mu_0\|_{\ell^2(\mu_m)\to\ell^q(\mu_0)}\|I\|_{\ell^q(\mu_0)\to\ell^\infty(\mu_0)}.$$

It follows from Corollary 4.4 that

$$\|K_{0,m}-\mu_0\|_{\ell^2(\mu_m)\to\ell^\infty(\mu_0)}\le\|I\|_{\ell^q(\mu_0)\to\ell^\infty(\mu_0)}\le\mu_0^{\#-1/q}.$$

Since $q=q(m)\ge\log(\mu_0^{\#-1})$ we get $\|K_{n-m,n}-\mu_n\|_{\ell^1(\mu_n)\to\ell^2(\mu_{n-m})}\le e$. $\square$

THEOREM 4.10. *Fix $c \in (1,\infty)$. Let V be a finite set equipped with a sequence of Markov kernels $(K_n)_1^\infty$. Assume that $(K_n)_1^\infty$ is c-stable with respect to a positive probability measure μ_0. For each i, set $\mu_0^i = \mu_0 K_i$ and $\mu_n^i = \mu_n K_i$. Let $\sigma_1(K_i, \mu_0^i)$ be the second largest singular value of the operator $K_i : \ell^2(\mu_0^i) \to \ell^2(\mu_0)$. Let $l(K_i^* K_i)$ be the logarithmic Sobolev constant of the operator $K_i^* K_i : \ell^2(\mu_0^i) \to \ell^2(\mu_0^i)$ where K_i^* is the adjoint of $K_i : \ell^2(\mu_0^i) \to \ell^2(\mu_0)$. Let $l(K_i K_i^*)$ be the logarithmic Sobolev constant of the operator $K_i K_i^* : \ell^2(\mu_n) \to \ell^2(\mu_n)$ where K_i^* is the adjoint of $K_i : \ell^2(\mu_n^i) \to \ell^2(\mu_n)$. If $\mu_i^\# = \min_x\{\mu_i(x)\}$ and*

$$\tilde{m}_0^\# = \min\left\{ t \in \mathbb{N} : \sum_{i=1}^{t} \log(1 + c^{-2} l(K_i^* K_i)) \geq \log\log(\mu_0^{\# -1/2}) \right\},$$

$$\tilde{m}_n^\# = \min\left\{ t \in \mathbb{N} : \sum_{i=n-t}^{n} \log(1 + c^{-6} l(K_i K_i^*)) \geq \log\log(\mu_n^{\# -1/2}) \right\},$$

then for any $n \geq 2\tilde{m}$

$$\max_{x,y}\left\{ \left| \frac{K_{0,n}(x,y)}{\mu_n(y)} - 1 \right| \right\} \leq e^2 \prod_{i=\tilde{m}}^{n-\tilde{m}} (1 - c^{-2}(1 - \sigma_1(K_i, \mu_0^i)^2))^{1/2},$$

where $\tilde{m} = \max\{\tilde{m}_0^\#, \tilde{m}_n^\#\}$.

PROOF. Note that $\mu_i/\mu_0^i \in [c^{-1}, c]$ and $\mu_i/\mu_n^i \in [c^{-2}, c^2]$. Let P_i and $\check{P}_i$ be the Markov kernels described in Theorem 4.8 with kernels

$$P_i(x,y) = \frac{1}{\mu_i(x)} \sum_z \mu_{i-1}(z) K_i(z,x) K_i(z,y), \tag{4.3}$$

$$\check{P}_i(x,y) = \sum_z \frac{\mu_{i-1}(y)}{\mu_i(z)} K_i(x,z) K_i(y,z). \tag{4.4}$$

Similar reasoning to that of Theorem 4.6 gives

$$l(P_i) \geq c^{-2} l(K_i^* K_i) \quad \text{and} \quad 1 - \sigma(P_i)^2 \geq c^{-2}(1 - \sigma(K_i, \mu_0^i)^2),$$

where K_i^* above is the adjoint of $K_i : \ell^2(\mu_0^i) \to \ell^2(\mu_0)$. This implies that $\tilde{m}_0^\# \geq m_0^\#$ where $m_0^\#$ is defined in Theorem 4.8.

In the case of $\check{P}_i$, equation (4.4) gives

$$\check{P}_i \geq c^{-4} \sum_z \frac{\mu_n(y)}{\mu_n^i(z)} K_i(x,z) K_i(y,z) = c^{-4} K_i K_i^*(x,y),$$

where K_i^* is the adjoint of the operator $K_i : \ell^2(\mu_n^i) \to \ell^2(\mu_n)$. A simple comparison argument yields

$$l(\check{P}_i) \geq c^{-6} l(K_i K_i^*)$$

and so $\tilde{m}_n^{\#} \geq m_n^{\#}$ where $m_n^{\#}$ is defined in Theorem 4.8. The desired result now follows from Theorem 4.8. $\square$

The next theorem gives us similar results when we have c-stability for a family of kernels.

THEOREM 4.11. *Fix $c \in (1, \infty)$. Let $\mathcal{Q}$ be a family of irreducible aperiodic Markov kernels on V. Assume that $\mathcal{Q}$ is c-stable with respect to some positive probability measure μ_0. Let $(K_n)_1^\infty$ be a sequence with $K_i \in \mathcal{Q}$ for all $i \geq 1$. Let π_i be the invariant measure of K_i and $\sigma_1(K_i)$ the second largest singular value of the operator K_i acting on $\ell^2(\pi_i)$. Let $l(K_i^* K_i)$ and $l(K_i K_i^*)$ be the logarithmic Sobolev constants of the operators $K_i^* K_i$ and $K_i K_i^*$ where K_i^* is the adjoint of $K_i : \ell^2(\pi_i) \to \ell^2(\pi_i)$. If $\mu_i^{\#} = \min_x\{\mu_i(x)\}$ and*

$$\tilde{m}_0^{\#} = \min\left\{t \in \mathbb{N} : \sum_{i=1}^{t} \log(1 + c^{-4} l(K_i^* K_i)) \geq \log\log(\mu_0^{\# -1/2})\right\},$$

$$\tilde{m}_n^{\#} = \min\left\{t \in \mathbb{N} : \sum_{i=n-t}^{n} \log(1 + c^{-6} l(K_i K_i^*)) \geq \log\log(\mu_n^{\# -1/2})\right\},$$

then for any $n \geq 2\tilde{m}$

$$\max_{x,y}\left\{\left|\frac{K_{0,n}(x,y)}{\mu_n(y)} - 1\right|\right\} \leq e^2 \prod_{i=\tilde{m}}^{n-\tilde{m}} (1 - c^{-4}(1 - \sigma_1(K_i)^2))^{1/2},$$

where $\tilde{m} = \max\{\tilde{m}_0^{\#}, \tilde{m}_n^{\#}\}$.

PROOF. First, note that $\mu_i/\pi_i \in [c^{-2}, c^2]$. Equation (4.3) implies that

$$l(P_i) \geq c^{-4} l(K_i^* K_i) \quad \text{and} \quad 1 - \sigma(K_i, \mu_i)^2 \geq c^{-4}(1 - \sigma(K_i)^2).$$

To bound $l(\check{P}_i)$, we use (4.4) to get that for all $x, y \in V$

$$\check{P}_i(x,y) \geq c^{-4} K_i K_i^*(x,y).$$

This implies that $l(\check{P}_i) \geq c^{-6} l(K_i K_i^*)$. It follows that $\tilde{m} \geq m$ where m is defined in Theorem 4.8. Applying Theorem 4.8 now gives us the desired result. $\square$

4.4. *An inhomogeneous walk on the hypercube.* Denote by $V = \{0,1\}^{2N}$ the $2N$-dimensional hypercube, we say that $x, y \in V$ are neighbors, or $x \sim y$ if

$$\sum_{i=1}^{N} |x_i - y_i| = 1,$$

where x_i is the ith coordinate of $x \in V$. The simple random walk on V is driven by the kernel

$$K(x,y) = \begin{cases} \dfrac{1}{2N}, & \text{if } x \sim y, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to check that μ, the uniform measure on V, is stationary for K.

Fix $\varepsilon \in (0,1)$ and consider the following perturbed version of K.

$$K_\varepsilon(x,y) = \begin{cases} \dfrac{1}{2N}, & \text{if } x \sim y \text{ and } |x| \neq N, \\ \dfrac{1+\varepsilon}{2N}, & \text{if } x \sim y \text{ and } |x| = N, y = |N|+1, \\ \dfrac{1-\varepsilon}{2N}, & \text{if } x \sim y \text{ and } |x| = N, y = |N|-1, \\ 0, & \text{otherwise.} \end{cases}$$

For $\varepsilon \in (0,1)$, set

$$\mathcal{Q}(\varepsilon) = \{K_\delta : \delta \in [-\varepsilon, \varepsilon]\}.$$

The example of time inhomogeneous Markov chains associated to $Q(\varepsilon)$ above is related to the binomial example in [32]. See Remark 4.17 below.

We shall show that $\mathcal{Q}(\varepsilon)$ is c-stable. First, consider the following definition.

DEFINITION 4.12. Let $\mathcal{S}_{2N}$ be the set of probability measures on $V = \{0, 1\}^{2N}$ that satisfy the following three properties:

(1) For all $x \in V$ with $|x| = N$ we have $\nu(x) = \frac{1}{4^N}$.
(2) For all $i \in \{-N, \dots, -1, 1, \dots, N\}$ there exists constants $a_{\nu,i}$ such that $a_{\nu,i} = -a_{\nu,-i}$ and for any x with $|x| = N + i$ we have

$$\nu(x) = \frac{1}{4^N} + a_{\nu,i}.$$

(3) For all $i \in \{-N, \dots, -1, 1, \dots, N\}$ we have $|a_{\nu,i}| \leq \varepsilon/4^N$.

CLAIM 4.13. *Let ν be in $\mathcal{S}_{2N}$ defined above, then for any $K \in \mathcal{Q}(\varepsilon)$ we have that $\nu K \in \mathcal{S}_{2N}$.*

PROOF. Let $\nu \in \mathcal{S}_{2N}$ and $Q \in \mathcal{Q}(\varepsilon)$, then $Q = K_\delta$ for some $\delta \in [-\varepsilon, \varepsilon]$. We will check each condition needed for νQ to be in $\mathcal{S}_{2N}$ separately.

(1) For any x with $|x| = N$ we have that $\nu Q(x) = \nu K(x)$. The desired result now follows from the definition of $\mathcal{S}_{2N}$.

(2) For i such that $|i| \notin \{1, N\}$, consider an element x such that $|x| = N + i$. Then

$$\begin{aligned}
\nu Q(x) &= \sum_{\substack{y\sim x\\ |y|=|x|+1}} \nu(y)Q(y,x) + \sum_{\substack{y\sim x\\ |y|=|x|-1}} \nu(y)Q(y,x)\\
&= \left(\frac{1}{2N}\right)\Bigg(\sum_{\substack{y\sim x\\ |y|=|x|+1}} \nu(y) + \sum_{\substack{y\sim x\\ |y|=|x|-1}} \nu(y)\Bigg)\\
&= \left(\frac{1}{2N}\right)\left[\left(\frac{1}{4^N} + a_{\nu,i+1}\right)|x| + \left(\frac{1}{4^N} + a_{\nu,i-1}\right)(2N - |x|)\right]\\
&= \frac{1}{4^N} + \frac{1}{2N}(a_{\nu,i+1}|x| + a_{\nu,i-1}(2N - |x|)).
\end{aligned}$$

A similar computation as above yields that for an element x with $|x| = N - i$ we have

$$\nu Q(x) = \frac{1}{4^N} - \frac{1}{2N}(a_{\nu,i+1}|x| + a_{\nu,i-1}(2N - |x|)).$$

When $i = N$, and x is such that $|x| = N + i = 2N$, we have

$$\begin{aligned}
\nu Q(x) &= \sum_{\substack{y\sim x\\ |y|=2N-1}} \nu(y)Q(y,x)\\
&= \left(\frac{1}{2N}\right)\left(\frac{1}{4^N} + a_{\nu,N-1}\right)(2N)\\
&= \frac{1}{4^N} + a_{\nu,N-1}.
\end{aligned}$$

When $i = -N$, and x is such that $|x| = N - i = 0$ we get $\nu Q(x) = \frac{1}{4^N} - a_{\nu,N-1}$ as desired.

Finally, we check that cases for elements x with $|x| = N \pm 1$. Consider an x such that $|x| = N - 1$, then

$$\begin{aligned}
\nu Q(x) &= \sum_{\substack{y\sim x\\ |y|=N-2}} \nu(y)Q(y,x) + \sum_{\substack{y\sim x\\ |y|=N}} \nu(y)Q(y,x)\\
&= \left(\frac{1}{2N}\right)\left(\frac{1}{4^N} - a_{\nu,2}\right)(N-1) + \left(\frac{1-\delta}{2N}\right)\left(\frac{1}{4^N}\right)(N+1)\\
&= \frac{1}{4^N} - \frac{1}{2N}\left(a_{\nu,2}(N-1) + \frac{\delta(N+1)}{4^N}\right).
\end{aligned}$$

When $|x| = N+1$, then

$$\begin{aligned}
\nu Q(x) &= \sum_{\substack{y\sim x\\ |y|=N}} \nu(y)Q(y,x) + \sum_{\substack{y\sim x\\ |y|=N+2}} \nu(y)Q(y,x)\\
&= \left(\frac{1+\delta}{2N}\right)\left(\frac{1}{4^N}\right)(N+1) + \left(\frac{1}{2N}\right)\left(\frac{1}{4^N}+a_{\nu,2}\right)(N-1)\\
&= \frac{1}{4^N} + \frac{1}{2N}\left(a_{\nu,2}(N-1) + \frac{\delta(N+1)}{4^N}\right)
\end{aligned}$$

as desired. We can now concluded that $a_{\nu Q,i} = -a_{\nu Q,-i}$.

(3) From the calculations in part (2), we know that for x with $x = N+i$ and $|i| \notin \{1, N\}$ and $|i| = N$ we have

$$\nu Q(x) = \frac{1}{4^N} + \frac{1}{2N}(a_{\nu,i+1}|x| + a_{\nu,i-1}(2N-|x|))$$

and

$$\nu Q(x) = \frac{1}{4^N} + a_{\nu,N-1},$$

respectively. It follows from the fact that for all i, $|a_{\nu,i}| \le \varepsilon/4^N$ that for both cases above $|a_{\nu Q,i}| \le \varepsilon/4^N$. When $|i| = 1$, we have that for x with $|x| = N+i = N\pm 1$

$$\begin{aligned}
\nu Q(x) &\le \frac{1}{4^N} + \frac{1}{2N}\left(a_{\nu,2}(N-1) + \frac{\varepsilon(N+1)}{4^N}\right)\\
&\le \frac{1}{4^N} + \frac{1}{2N}\left(\frac{\varepsilon(N-1)}{4^N} + \frac{\varepsilon(N+1)}{4^N}\right)\\
&= \frac{1+\varepsilon}{4^N}.
\end{aligned}$$

A similar calculation yields $\nu Q(x) \ge \frac{1-\varepsilon}{4^N}$. The proof now follows from the fact that $a_{\nu Q,i} = -a_{\nu Q,-i}$. $\square$

Claim 4.14. *The set $\mathcal{Q}(\varepsilon)$ is $\frac{1+\varepsilon}{1-\varepsilon}$-stable with respect to any measure in $\mathcal{S}_{2N}$.*

Proof. Let $\mu_0 \in \mathcal{S}_{2N}$. Let $(K_i)_1^\infty$ be any sequence of kernels such that $K_i \in \mathcal{Q}(\varepsilon)$ for all $i \ge 1$. Let $\mu_n = \mu_0 K_{0,n}$, then by Claim 4.13 we have that $\mu_n \in \mathcal{S}_{2N}$ and so for any $x \in V$

$$\frac{1-\varepsilon}{1+\varepsilon} \le \frac{\mu_n(x)}{\mu_0(x)} \le \frac{1+\varepsilon}{1-\varepsilon}. \qquad \square$$

The kernels $K_\delta \in \mathcal{Q}(\varepsilon)$ drive periodic chains that will alternate between points with an even number of 1's and odd number of 1's. So we will study following random walk driven by the kernel

$$Q_\delta = \tfrac{1}{2}(I + K_\delta),$$

where I is the identity. Set

$$\overline{\mathcal{Q}}(\varepsilon) = \{Q_\delta : \delta \in [-\varepsilon, \varepsilon]\}.$$

Claim 4.15. *Let $(K_i)_1^\infty$ be a sequence of Markov kernels such that $K_i \in \overline{\mathcal{Q}}(\varepsilon)$ for all $i \geq 1$. Let $\mu_0 \in \mathcal{S}_{2N}$ be a positive measure, and let $\mu_n = \mu_0 K_{0,n}$. Set $P_i = K_i^* K_i : \ell^2(\mu_i) \to \ell^2(\mu_i)$ where K_i^* is the adjoint of $K_i : \ell^2(\mu_i) \to \ell^2(\mu_{i-1})$. Let $\sigma_1(K_i, \mu_i)$ and be the second largest singular value of $K_i : \ell^2(\mu_i) \to \ell^2(\mu_{i-1})$. Let $l(P_i)$ be logarithmic Sobolev constant of P_i. Then*

$$\sigma_1(K_i, \mu_i) \leq 1 - C(\varepsilon)\frac{1}{2N} \quad \textit{and} \quad l(P_i) \geq \frac{C(\varepsilon)}{4N},$$

where $C(\varepsilon) = (1+\varepsilon)^{-2}(1-\varepsilon)^4$.

Proof. Let $Q = 2^{-1}(I + K_0)$ and u be the uniform measure on $\{0,1\}^{2N}$. Let $P_i(x,y) = K_i^* K_i : \ell^2(\mu_i) \to \ell^2(\mu_i)$. Using the $\frac{1+\varepsilon}{1-\varepsilon}$-stability of the sequence $(\mu_n)_0^\infty$, we get that

$$\begin{aligned}
\mu_i(x) P_i(x,y) &= \sum_z \mu_{i-1}(z) K_i(z,x) K_i(z,y) \\
&\geq \frac{1-\varepsilon}{1+\varepsilon} \frac{u(x)}{u(x)} \sum_z u(z) K_i(z,x) K_i(z,y) \\
&\geq \frac{(1-\varepsilon)^3}{1+\varepsilon} \frac{u(x)}{u(x)} \sum_z u(z) Q(z,x) Q(z,y) \\
&\geq \frac{(1-\varepsilon)^3}{1+\varepsilon} u(x) Q^{(2)}(x,y).
\end{aligned}$$

A simple comparison yields

$$\mathcal{E}_{P_i,\mu_i}(f,f) \geq (1-\varepsilon)^3 (1+\varepsilon)^{-1} \mathcal{E}_{Q^{(2)},u}(f,f).$$

Further comparison gives that

$$1 - \sigma_1(K_i, \mu_i) \geq C(\varepsilon)(1 - \sigma_1(Q)), \tag{4.5}$$

$$l(P_i) \geq C(\varepsilon) l(Q^{(2)}). \tag{4.6}$$

It is well known that for $K_0:\ell^2(u)\to\ell^2(u)$ (the simple random walk) we have $2l(K_0)=1-\sigma_1(K_0)=1/N$. This implies that $\sigma_1(Q)=1-1/2N$. The singular value inequality in Claim 4.15 now follows from (4.5). For the rest of the proof, we note that Lemma 2.5 of [11] tells us that $\mathcal{E}_{Q^{(2)},u}(f,f)\geq \mathcal{E}_{Q,u}(f,f)$, and so we get $l(Q^{(2)})\geq l(Q)$. The logarithmic Sobolev inequality now follows from (4.6) and the fact that $l(Q)=1/4N$. $\square$

By applying Theorem 4.5 and Claim 4.15, we get the following theorem.

THEOREM 4.16. *For any $\varepsilon\in(0,1)$ there exists a constant $D(\varepsilon)$ such that the total variation merging time of the sequence $(K_i)_1^\infty$ with $K_i\in\overline{\mathcal{Q}}(\varepsilon)$ for all $i\in\{1,2,\dots\}$ is bounded by*

$$T_{\mathrm{TV}}(\eta)\leq D(\varepsilon)N(\log N+\log_+ 1/\eta).$$

Moreover, we can chose $D(\varepsilon)$ such that

$$\forall n\geq D(\varepsilon)N(\log N+\log_+ 1/\eta)\qquad \max_{x,y,z\in V}\left\{\left|\frac{K_{0,n}(x,z)}{K_{0,n}(y,z)}-1\right|\right\}\leq\eta.$$

We note that the relative-sup merging time bound is obtained with the same arguments as those used at the end of the proof of Theorem 2.4.

REMARK 4.17. The theorem above is closely related to the example in Section 5.2 of [32] which studies a time inhomogeneous chain on $\{-N,\dots,N\}$ resulting from perturbations of a birth and death chain with binomial stationary distribution. Both [32] and Theorem 4.16 give the correct upper bound on the merging time yet [32] requires knowledge about the entire spectrum of the operators driving the chain while the theorem above uses logarithmic Sobolev techniques.

4.5. *Modified logarithmic Sobolev inequalities and entropy.* Let ν and $\mu>0$ be two probability measures on V. Define the relative entropy between μ and ν as

$$\mathrm{Ent}_\mu(\nu)=\sum_{x\in V}\mu(x)\log\left(\frac{\mu(x)}{\nu(x)}\right).$$

It is well known that $\sqrt{2}\|\mu-\nu\|_{\mathrm{TV}}\leq\sqrt{\mathrm{Ent}_\nu(\mu)}$. Let $(K_n)_0^\infty$ be a sequence of Markov kernels on V, μ_0 be some initial distribution on V and $\mu_n=\mu_0K_{0,n}$. It follows by the triangle inequality that for any $x,y\in V$

$$\|K_{0,n}(x,\cdot)-K_{0,n}(y,\cdot)\|_{\mathrm{TV}}\leq\sqrt{2}\max_{x\in V}\sqrt{\mathrm{Ent}_{\mu_n}(K_{0,n}(x,\cdot))}.$$

Let $\alpha = \alpha(K, \nu)$ be the largest constant such that for any probability measure μ

$$\operatorname{Ent}_{\nu K}(\mu K) \leq (1-\alpha)\operatorname{Ent}_{\nu}(\mu).$$

Let $\mu' = \mu K$ and $K^* : \ell^2(\mu) \to \ell^2(\mu')$ be the adjoint of $K : \ell^2(\mu') \to \ell^2(\mu)$. Set

$$\check{P} = KK^* : \ell^2(\mu) \to \ell^2(\mu).$$

In [7], the contraction constant α is related to the so-called modified logarithmic Sobolev constant

$$l'(\check{P}) = \inf\left\{\frac{\mathcal{E}_{\mu,\check{P}}(f^2, \log(f^2))}{\mathcal{L}(f^2, \mu)} : \mathcal{L}(f^2, \mu) \neq 0, f \neq \text{constant}\right\}.$$

Proposition 4.18 ([7], Proposition 5.1). *There exists a universal constant $0 < \rho < 1$ such that for any Markov kernel K and any probability measure μ,*

$$\rho l'(\check{P}) \leq \alpha(K, \mu) \leq l'(\check{P}),$$

where $\check{P} = KK^*$ and K^* is the adjoint of the operator $K : \ell^2(\mu') \to \ell^2(\mu)$, $\mu' = \mu K$.

Proposition 4.19. *Referring to the proposition above,*

$$\rho \geq \log 2\left(\frac{1-\log 2}{2}\right).$$

Proof. The proof of Proposition 5.1 in [7] uses the fact that there exists some $0 < \tilde{\rho} < 1$ such that for all $x \in [-1, \infty)$

$$0 \leq \varphi(x) \leq \tilde{\rho}^{-1}\varphi(x/2),$$

where

$$\varphi(x) = (1+x)\log(1+x) - x.$$

Let $f(x) = \varphi(x) - (2/(1-\log 2))\varphi(x/2)$. We will show that for all $x \in [-1, \infty)$ then $f(x) \leq 0$. By differentiating f we get

$$f'(x) = \log(1+x) - \left(\frac{1}{1-\log 2}\right)\log\left(\frac{2+x}{2}\right)$$

and

$$f'''(x) = \frac{4\log 2(1+x) + x^2\log 2 - 3 - 2x}{(1+x)^2(2+x)^2(1-\log 2)}.$$

In particular, for $x \in [-1,0]$ we have $f'''(x) \le 0$. This along with the fact that

$$f'(-0.9) \le 0, \qquad f'(-0.1) > 0 \quad \text{and} \quad f'(0) = 0$$

implies that there exists only one $z \in (-1,0)$ such that $f'(z) = 0$. It follows that f is decreasing on $[-1,z]$ and f is increasing on $[z,0]$. Since $f(-1) = f(0) = 0$, then for $x \in [-1,0]$ we have that $f(x) \le 0$.

For $x \in [0,\infty)$, we note that

$$f''(x) = \frac{1}{1+x} - \frac{1}{(1-\log 2)(2+x)} \le 0,$$

which implies that $f'(x) \le f'(0) = 0$. The fact that $f(x) \le f(0) = 0$ implies $\tilde{\rho} = 2/(1-\log 2)$. The desired result follows from the fact that the proof of Proposition 5.1 in [7] shows that

$$\alpha(K,\mu) \ge \tilde{\rho} \log(2) l'(\check{P}). \qquad \square$$

The results in [7] allow us to study merging via logarithmic Sobolev constants.

PROPOSITION 4.20. *Let V be a finite state space equipped with a sequence of Markov kernels $(K_n)_1^\infty$ and an initial distribution μ_0. Let $\mu_n = \mu_0 K_{0,n}$ and $\check{P}_i = K_i K_i^* : \ell^2(\mu_{i-1}) \to \ell^2(\mu_{i-1})$ where K_i^* is the adjoint of $K_i : \ell^2(\mu_i) \to \ell^2(\mu_{i-1})$. Set $\mu_0^* = \min_x \mu_0(x)$ then for any $x, y \in V$*

$$\|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} \le \sqrt{2} \log\left(\frac{1}{\mu_0^*}\right)^{1/2} \prod_{i=1}^n (1 - \rho l'(\check{P}_i))^{1/2},$$

where ρ is given in Propositions *4.18* and *4.19*.

PROOF. We note that for any $x, y \in V$

$$\|K_{0,n}(x,\cdot) - K_{0,n}(y,\cdot)\|_{\mathrm{TV}} \le \sqrt{2} \max_{x,y} \sqrt{\mathrm{Ent}_{\mu_n}(K_{0,n}(x,\cdot))}.$$

Proposition 5.1 in [7] gives that

$$\mathrm{Ent}_{\mu_n}(K_{0,n}(x,\cdot)) \le \mathrm{Ent}_{\mu_0}(\delta_x) \prod_{i=1}^n (1 - \rho l'(\check{P}_i)).$$

The desired result now follows from the fact that

$$\mathrm{Ent}_{\mu_0}(\delta_x) = \log\left(\frac{1}{\mu_0(x)}\right) \le \log\left(\frac{1}{\mu_0^*}\right). \qquad \square$$

4.6. *Biased shuffles.* In this section, we present two examples where the modified logarithmic Sobolev inequality technique yields the correct merging time while the regular logarithmic Sobolev inequality technique does not. Let $V_n = S_n$ be the symmetric group equipped with the uniform probability measure u. Let $\tilde{Q}_i$ be the the kernel of transpose i with random, that is,

$$\tilde{Q}_i(x,y) = \begin{cases} 1/n, & \text{if } x^{-1}y = (i,j) \text{ for } j \in [1,n], \\ 0, & \text{otherwise.} \end{cases}$$

Let $Q_i = 2^{-1}(I + \tilde{Q}_i)$ be the associated lazy chain. It is known that the lazy chain has a mixing time of of $2n \log n$. More precisely,

$$t \geq 2n(\log n + c) \quad \Rightarrow \quad \max_{x,y} \left\{ \frac{Q_i^t(x,y)}{u(y)} - 1 \right\} \leq 2e^{-2c} \qquad \forall x \in S_n.$$

See, for example, [31]. The results of [18] show that the modified logarithmic Sobolev constant for Q_i is bounded by

$$\frac{1}{n-1} \geq l'(Q_i) \geq \frac{1}{4(n-1)}.$$

Set $\mathcal{Q} = \{Q_i, i = 1, \ldots, n\}$. Since all Q_i are reversible with respect to the uniform distribution u, the set $\mathcal{Q}$ is 1-stable with respect to u. Using the methods of [30] (see also [17, 24]), one can prove that for any sequence $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}$ for all $i \geq 1$ we have

$$t \geq 2n(\log n + c) \quad \Rightarrow \quad \max_{x,y} \left\{ \frac{K_{0,n}(x,y)}{u(y)} - 1 \right\} \leq 2e^{-2c} \qquad \forall x \in S_n.$$

The inequality above is due to the fact that the Q_i are driven by probability measures so the ℓ^2 distance bounds the ℓ^∞ distance and the eigenvectors in Theorem 3.2 of [32] drop out to give

$$d_2(K_{0,t}(x,\cdot), u)^2 \leq \sum_{i=1}^{n!-1} \prod_{j=1}^{t} \sigma_i(K_j)^2. \tag{4.7}$$

One can then group the singular values in the equality above since the Q_i's are all are images of each other under some inner automorphism of S_n which implies $\sigma_j(Q_i) = \sigma_j(Q_k)$ for all i, j, k. For a more detailed discussion, see [30].

We now consider two variants of this example that cannot be treated using the singular values techniques of [17, 30, 32] or the logarithmic Sobolev inequality technique of Sections 4.1–4.4 but where the modified logarithmic Sobolev inequality does yield a successful analysis. This technique can be applied to the two examples in this section because of the following three reasons:

(1) any sequence $(K_i)_1^\infty$ of interest can be shown to be c-stable with respect to some well chosen initial distribution;

(2) all the kernels K_i driving the time inhomogeneous process are directly comparable to the Q_i's and,
(3) due to (1) and the laziness of the Q_i's we can successfully estimate the modified logarithmic Sobolev constants $l'(Q_iQ_i^*) = l'(Q_i^{(2)})$ to be of order $1/n$.

4.6.1. *Symmetric perturbations in S_n.* For the first variant, fix $\varepsilon \in (0,1)$ and consider the set $\mathcal{Q}^{\#}(\varepsilon)$ of all Markov kernels K on S_n such that:

(a) $K(x,y) = K(y,x)$ (symmetry) and
(b) $\forall x, y$ we have $(1-\varepsilon)Q_i(x,y) \leq K(x,y) \leq (1+\varepsilon)Q_i(x,y)$ for some $i \in \{1,\ldots,n\}$.

Hence, $\mathcal{Q}^{\#}(\varepsilon)$ is the set of all symmetric edge perturbations of kernels in $\mathcal{Q}$. As we require symmetry, the uniform distribution is invariant for all the kernels in $\mathcal{Q}^{\#}(\varepsilon)$. Now, what can be said of the merging properties of sequences $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}^{\#}(\varepsilon)$? Unlike $\mathcal{Q}$, the kernels in $\mathcal{Q}^{\#}(\varepsilon)$ are not invariant under left multiplication in S_n. So the eigenvectors of Theorem 3.2 in [32] do not drop out, and we only get

$$d_2(K_{0,t}(x,\cdot),u)^2 \leq n! \prod_{i=1}^{t} \sigma_1(K_i,u)^2.$$

Singular value comparison yields $\sigma_1(K_i,u) \leq 1-(1-\varepsilon)/(2n)$ which gives

$$t \geq (1-\varepsilon)^{-1} n(n\log n + 2c) \quad \Rightarrow \quad d_2(K_{0,t}(x,\cdot),u) \leq e^{-c} \qquad \forall x \in S_n.$$

This indicates merging after order $n^2 \log n$ steps instead of the expected order $n \log n$ steps. For any sequence $(K_i)_1^\infty$ with $K_i \in \mathcal{Q}^{\#}(\varepsilon)$ for all $i \geq 1$ set $\check{P}_i = K_iK_i^*$ where K_i^* is the adjoint of the operator $K_i : \ell^2(u) \to \ell^2(u)$. A simple comparison argument gives

$$\check{P}_i(x,y) \geq (1-\varepsilon)^2 Q_j^2(x,y)$$

for some $j \in [1,n]$. Further comparison yields $l'(\check{P}_i) \geq (1-\varepsilon)^2 l'(Q_j^2)$. Lemma 2.5 of [11] implies that $l'(Q_j^2) \geq l'(Q_j)$ so $l'(\check{P}_i)$ is of order at least $1/n$. Hence, there exists some constant $C(\varepsilon)$ independent of n such that

$$\|K_{0,t}(x,\cdot) - K_{0,t}(y,\cdot)\|_{\mathrm{TV}} \leq \sqrt{2\log n!}(1 - C(\varepsilon)/n)^{t/2}.$$

In particular, for some constant $D(\varepsilon)$ we get $T_{\mathrm{TV}}(\eta) \leq D(\varepsilon)n(\log n + \log_+ 1/\eta)$. To obtain a result for the relative-sup norm, one can use the (nonmodified) logarithmic Sobolev technique as the modified logarithmic Sobolev technique only gives bounds in total variation. It is known that the logarithmic Sobolev constant for top to random is of order $1/(n\log n)$, see

[20], leading to results that are off by a factor of $\log n$. This technique yields the best available result,

$$t \geq C(\varepsilon)n((\log n)^2 + c) \quad \Rightarrow \quad \max_{x,y,z}\left\{\left|\frac{K_{0,t}(x,z)}{K_{0,t}(y,z)} - 1\right|\right\} \leq e^{-c}.$$

4.6.2. *Sticky permutations.* We now consider a second variation on the transpose cyclic to random example. Let $\rho \in S_n$, $\delta \in (0, 1 - Q_1(\rho, \rho))$ and consider the Markov kernel

$$K(x,y) = \begin{cases} Q_1(x,y), & \text{if } x \neq \rho, \\ Q_1(x,y) + \delta, & \text{if } x = y = \rho, \\ Q_1(x,y) - \delta/(n-1), & \text{if } x = \rho \text{ and } x^{-1}y = (1,j) \text{ for } j \in [2,n]. \end{cases}$$

In words, K is obtained from Q_1 by adding extra holding probability at ρ, making ρ "sticky." Next, if σ is the cycle $(1, \ldots, n)$, let

$$K_i(x,y) = K(\sigma^{i-1}x\sigma^{-i+1}, \sigma^{i-1}y\sigma^{-i+1}).$$

In words, K_i is Q_i with some added holding at $\rho_i = \sigma^{-i+1}\rho\sigma^{i-1}$.

We would like to consider the merging properties of the sequence $(K_i)_1^\infty$. Unlike the previous example, the uniform probability is not invariant under K_i. However, this type of construction is considered in [33].

Let

$$\varepsilon = \frac{\delta}{\sum_{z \neq \rho} Q_1(\rho, z)}$$

so that $K(x,y) \geq (1-\varepsilon)Q_1(x,y)$. It is proved that $(K_i)_1^\infty$ is $(1-\varepsilon)^{-1}$-stable with respect to the probability measure $\mu_0 = \tilde{\pi}$, where $\tilde{\pi}$ is the invariant probability measure of the Markov kernel $\tilde{K}(x,y) = K(x, \sigma^{-1}y\sigma)$. From the analysis in [33], Section 5, one can see that

$$(1-\varepsilon)u \leq \tilde{\pi} \leq (1-\varepsilon)^{-1}u.$$

Applying the singular value techniques used in Section 5 of [33] would give us an upper bound on the relative sup merging time of order $n^2 \log n$.

Set $\check{P}_i = K_i K_i^* : \ell^2(\mu_{i-1}) \to \ell^2(\mu_{i-1})$ where K_i^* is the adjoint of the operator $K_i : \ell^2(\mu_i) \to \ell^2(\mu_{i-1})$. Since $K_i(x,y) \geq (1-\varepsilon)Q_i(x,y)$, for $x \neq y$ we can write

$$\check{P}_i(x,y) = \sum_z K_i(x,z)K_i(y,z)\mu_{i-1}(y)\mu_i(z)^{-1} \geq (1-\varepsilon)^4 Q_i^2(x,y).$$

It follows by comparison that $l'(\check{P}_i) \geq (1-\varepsilon)^5 l'(Q_i^2)$. We can successfully estimate $l'(Q_i^2)$ due to Lemma 2.5 of [11] which implies $l'(Q_i^2) \geq l'(Q_i)$. So

we have that $l'(\check{P}_i)$ is at least $(1-\varepsilon)^5/(4(n-1))$. Proposition 4.20 gives us that

$$\|K_{0,t}(x,\cdot)-K_{0,t}(y,\cdot)\|_{\mathrm{TV}} \le \sqrt{2}\log\left(\frac{n!}{1-\varepsilon}\right)^{1/2}\left(1-\frac{\rho(1-\varepsilon)^5}{4(n-1)}\right)^{t/2},$$

where ρ is as in Proposition 4.19. So for some constant $D = D(\varepsilon)$, we get

$$T_{\mathrm{TV}}(\eta) \le Dn(\log n + \log_+(1/\eta)).$$

REFERENCES

[1] Aldous, D. (1983). Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on Probability, XVII. Lecture Notes in Math.* **986** 243–297. Springer, Berlin. MR770418

[2] Bobkov, S. G. and Tetali, P. (2006). Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.* **19** 289–336. MR2283379

[3] Condon, A. and Hernek, D. (1994). Random walks on colored graphs. *Random Structures Algorithms* **5** 285–303. MR1262980

[4] D'Aristotile, A., Diaconis, P. and Freedman, D. (1988). On merging of probabilities. *Sankhyā Ser. A* **50** 363–380. MR1065549

[5] Del Moral, P. and Guionnet, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincaré Probab. Statist.* **37** 155–194. MR1819122

[6] Del Moral, P. and Miclo, L. (2006). Dynamiques recuites de type Feynman–Kac: Résultats précis et conjectures. *ESAIM Probab. Stat.* **10** 76–140 (electronic). MR2218405

[7] Del Moral, P., Ledoux, M. and Miclo, L. (2003). On contraction properties of Markov kernels. *Probab. Theory Related Fields* **126** 395–420. MR1992499

[8] Diaconis, P. (1988). *Group Representations in Probability and Statistics. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **11**. Inst. Math. Statist., Hayward, CA. MR964069

[9] Diaconis, P. and Shahshahani, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** 159–179. MR626813

[10] Diaconis, P. and Saloff-Coste, L. (1993). Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3** 696–730. MR1233621

[11] Diaconis, P. and Saloff-Coste, L. (1996). Nash inequalities for finite Markov chains. *J. Theoret. Probab.* **9** 459–510. MR1385408

[12] Diaconis, P. and Saloff-Coste, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750. MR1410112

[13] Diaconis, P. and Saloff-Coste, L. (1998). What do we know about the Metropolis algorithm? *J. Comput. System Sci.* **57** 20–36. MR1649805

[14] Douc, R., Moulines, E. and Rosenthal, J. S. (2004). Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Ann. Appl. Probab.* **14** 1643–1665. MR2099647

[15] Fill, J. A. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.* **1** 62–87. MR1097464

[16] Fleischer, I. and Joffe, A. (1995). Ratio ergodicity for non-homogeneous Markov chains in general state spaces. *J. Theoret. Probab.* **8** 31–37. MR1308668

[17] Ganapathy, M. (2007). Robust mixing time. *Electron. J. Probab.* **12** 229–261.

[18] Goel, S. (2004). Modified logarithmic Sobolev inequalities for some models of random walk. *Stochastic Process. Appl.* **114** 51–79. MR2094147

[19] Iosifescu, M. (1980). *Finite Markov Processes and Their Applications*. Wiley, Chichester. MR587116

[20] Lee, T.-Y. and Yau, H.-T. (1998). Logarithmic Sobolev inequality for some models of random walks. *Ann. Probab.* **26** 1855–1873. MR1675008

[21] Miclo, L. (1997). Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies. In *Séminaire de Probabilités, XXXI. Lecture Notes in Math.* **1655** 136–167. Springer, Berlin. MR1478724

[22] Montenegro, R. and Tetali, P. (2006). Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.* **1** x+121. MR2341319

[23] Morris, B. and Peres, Y. (2005). Evolving sets, mixing and heat kernel bounds. *Probab. Theory Related Fields* **133** 245–266. MR2198701

[24] Mossel, E., Peres, Y. and Sinclair, A. (2004). Shuffling by semi-random transpositions. In *45th Symposium on Foundations of Comp. Sci.* Available at arXiv:math.PR/0404438.

[25] Neumann, M. and Schneider, H. (1999). The convergence of general products of matrices and the weak ergodicity of Markov chains. *Linear Algebra Appl.* **287** 307–314. MR1662874

[26] Păun, U. (2001). Ergodic theorems for finite Markov chains. *Math. Rep. (Bucur.)* **3** 383–390. MR1990903

[27] Rhodius, A. (1997). On the maximum of ergodicity coefficients, the Dobrushin ergodicity coefficient, and products of stochastic matrices. *Linear Algebra Appl.* **253** 141–154. MR1431171

[28] Saloff-Coste, L. (1997). Lectures on finite Markov chains. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1996). Lecture Notes in Math.* **1665** 301–413. Springer, Berlin. MR1490046

[29] Saloff-Coste, L. (1999). Simple examples of the use of Nash inequalities for finite Markov chains. In *Stochastic Geometry (Toulouse, 1996). Monographs on Statistics and Applied Probability* **80** 365–400. Chapman and Hall/CRC, Boca Raton, FL. MR1673142

[30] Saloff-Coste, L. and Zúñiga, J. (2007). Convergence of some time inhomogeneous Markov chains via spectral techniques. *Stochastic Process. Appl.* **117** 961–979. MR2340874

[31] Saloff-Coste, L. and Zúñiga, J. (2008). Refined estimates for some basic random walks on the symmetric and alternating groups. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 359–392. MR2461789

[32] Saloff-Coste, L. and Zúñiga, J. (2009). Merging for time inhomogeneous finite Markov chains. I. Singular values and stability. *Electron. J. Probab.* **14** 1456–1494. MR2519527

[33] Saloff-Coste, L. and Zúñiga, J. (2010). Time inhomogeneous Markov chains with wave like behavior. *Ann. Appl. Probab.* **20** 1831–1853.

[34] Saloff-Coste, L. and Zúñiga, J. (2009). Merging and stability for time inhomogeneous finite Markov chains. In *Proc. SPA Berlin*. To appear. Available at arXiv:1004.2296v1.

[35] Seneta, E. (1973). On strong ergodicity of inhomogeneous products of finite stochastic matrices. *Studia Math.* **46** 241–247. MR0332843

[36] Seneta, E. (2006). *Non-negative Matrices and Markov Chains*. Springer, New York. MR2209438

[37] Sonin, I. (1996). The asymptotic behaviour of a general finite nonhomogeneous Markov chain (the decomposition–separation theorem). In *Statistics, Probability and Game Theory. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **30** 337–346. Inst. Math. Statist., Hayward, CA. MR1481788

[38] Stein, E. M. and Wiess, G. (1971). *Introduction to Fourier Analysis in Euclidean Spaces*. Princeton Univ. Press, Princeton, NJ.

Department of Mathematics
Cornell University
Malott Hall
Ithaca, New York, 14853
USA

Department of Mathematics
Stanford University
Stanford, California, 94305
USA
E-mail: jzuniga@math.stanford.edu